Friend Me Your Ears: A Musical Approach to Human-Robot Relationships

Louis McCallum

Supervised by: Professor Peter McOwan, Dr Andrew McPherson

Thesis submitted in partial fulfilment
of the requirements of the University of London
for the Degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary University of London

2015

Abstract

A relationship is something that is necessarily built up over time, however, Human-Robot Interaction (HRI) trials are rarely extended beyond a single session. These studies are insufficient for examining multi-interaction scenarios, which will become commonplace if the robot is situated in a workplace or adopts a role that is part of a human's routine. Long term studies that have been executed often demonstrate a declining novelty effect. Music, however, provides an opportunity for affective engagement, shared creativity, and social activity. This being said, it is unlikely that a robot best equipped to build sustainable and meaningful relationships with humans will be one that can solely play music. In their day-to-day lives, most humans encounter machines and computer programs capable of executing impressively complex tasks to a high standard that may provide them with hours of engagement. In order to have anything that that could be classed as a social relationship, the human must have the sense that their interactions are taking place with another, a phenomenon known as social presence. In this thesis, we examine whether the addition of simulated social behaviours will improve a sense of believability or social presence, which, along with an engaging musical interaction, will allow us to move towards something that could be called a human-robot relationship. First, we conducted a large online survey to gain insight into relationships based in regular music activity. Using these results, we designed, constructed and programmed *Mortimer*, a robotic system capable of playing the drums and a responsive composition algorithm to best meet these aims. This robot was then used in a series of studies, one single session and two long-term, testing various simulated social behaviours to compliment the musical improvisation. These experiments and their results address the paucity of long-term studies both specifically in Social Robotics and in the broader HRI field, and provide a promising insight into a possible solution to generally poor outcomes in this area. This conclusion is based upon the model of a positive human-robot relationship and the methodological approach of automated behavioural metrics to evaluate robotic systems in this regard developed and detailed within the thesis.

Acknowledgements

I have received guidance and wisdom from many academics who have been kind enough to impart their considerable experience to me, without which I could not have reached this point, none more so than my supervisor, Peter McOwan. Further, Nick Collins set me off on this journey and Andrew McPherson and Geraint Wiggins gave valuable advice at key progression stages.

Shauna Concannon and Davy Smith took me into their home when I was in need of shelter and Joe Lyske has given me constant inspiration and support throughout, as have my mother and father.

My fellow students of the Media and Arts Technology programme have all provided advice and distraction in equal measure, especially Keir Williams, David Meckin, Vincent Akkermans, Matt Jarvis, Polly Barden, Toby Harris, Evan Morgan, Nicola Plant, Duncan Menzies, Ben Bengler and Artist-in-Residence Di Mainstone. Praise is also deserved by Pat Healey, Mark Sandler and the EPSRC (RCUK) for beginning this programme and Richard Kelly for keeping it afloat.

Many people have selflessly given up their time to provide me with the knowledge and skills necessary to complete this work. I'd like to thank Marcus Pearce, Daniel Schulman and Will Jolly for enlightening me to murky world of experimental statistics and Hamit Soyel, Andrew Robertson, Adam Stark, Julian Hough and Matt Purver for explaining their own excellent research.

I am indebted to the staff of QMUL for providing access to many crucial facilties of the institution such as space, equipment and funding. This includes David Wilkinson, Kok Ho Huen, Guilio Moro, Colin Powell, Laura King, Hayley Cork, Sharon Cording, Therese, the Research Degrees Office and the entire EECS IT dept.

Finally, I'd like to thank everyone that took part in studies and forwarded emails, without you it would just be me. In a room. With a robot.

This research is supported by the EPSRC as part of the Media and Arts Technology Doctoral Training Centre, EP/G03723X/2.

License

I, Louis McCallum, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third partys copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:
Date:

Table of Contents

List of Figures

List of Tables

Acronyms

ASR Automatic Speech Recognition. 69, 72, 89

BEV Behavioral Ecology View. 99

HRI Human-Robot Interaction. 1, 10–12, 21–23, 28, 30, 37–40, 43–46, 50–53, 56, 68, 69, 80, 87, 92, 94, 104, 131, 135–139, 141, 143

HRP Human-Robot Proxemics. 20

IMS Interactive Music System. 12, 28, 30–34, 36, 66, 95, 139

LEMUR The League of Electronic Musical Urban Robots. 35

LIREC Living with Robots and Interactive Companions. 20, 42

MIDI Musical Instrument Digital Interface. 30, 32–34, 75, 172

MIR Music Information Retrieval. 144

MPR Music Performance Robots. 34

NARS Negative Attitudes Towards Robots Scale. 38

NLG Natural Language Generation. 79

NRI Network of Relationships Index. 41, 55, 57, 107

NRI-SPV Network of Relationships Index Social Provisions Version. 12, 55, 57, 64, 87, 107, 112, 116, 124, 135

TPI Temple Presence Inventory. 50

TR Theatrical Robot. 40

VHRI Video based Human-Robot Interaction. 40

WoZ Wizard of Oz. 24, 40, 68, 137

Chapter 1

Introduction

1.1 Introduction

The field of Social Robotics focusses on either programming social awareness into robots completing tasks alongside humans or designing artificial assistants that leverage social skills to better achieve their goals. Within this, our interests lie with exploring how robot design and choice of interaction domain can allow for sustainable and meaningful human-robot relationships. A relationship is something that is necessarily built up over time, however, HRI trials are rarely extended beyond a single session. These studies are insufficient for examining multi-interaction scenarios, which will become commonplace if the robot is situated in a workplace or adopts a role that is part of a human's routine.

The subjects of the sporadic cases of long-term research include robotic household appliances [Sung et al., 2009], robotic pets [Fernaeus et al., 2010] and fully mobile anthropomorphic robots [Mavridis et al., 2011,Lee et al., 2012,Mitsunaga et al., 2006]. The interactions afforded by the first two are simply not engaging enough and result in a swiftly declining novelty effect. The latter often attempts to use either gesture or language as the basis for interactions and in most cases present, to greater or lesser extent, the frustration of adult participants at the small range of abilities and often limited scripts in comparison to the sophisticated social interactions they regularly partake in with other humans. These frustrations are often caused by a system's morphological design or choice of interaction domain indicating more advanced abilities than it can deliver and can also result in swiftly declining positive responses over time. Counter to all the above issues, any time music is played as part of an ensemble, you are guaranteed to have at least two people, in the majority of cases co-located, simultaneously focussing their attention towards the same task and cooperating towards a joint goal in a highly engaging and naturally progressive interaction.

Music provides an opportunity for affective engagement, shared creativity, and social activity. However, it is unlikely that a robot best equipped to build sustainable and meaningful relationships with humans will be one that can solely play music. In their day-to-day lives, most humans encounter machines and computer programs capable of executing impressively complex tasks to a high standard that may provide them with hours of engagement. In order to have anything that that could be classed as a social relationship, the human must have the sense that their interactions are taking place with another, a phenomenon known as social presence [Biocca et al., 2003]. This concept addresses similar aspects of a human's perception of a robot as the notion of believability, already prevalent in Sociable Robotics research [Breazeal, 2004, Aylett et al., 2011] and described as the amount to which a person can suspend their knowledge a that robot is inanimate and not actually in possession of the human faculties we attempt to make it display. This leads us to the purpose of this thesis, summarised by the following statement:

Thesis Statement In this thesis, we will examine whether the addition of simulated social behaviours will improve a sense of believability or social presence, which, along with an engaging musical interaction, will allow us to move towards something that could be called a human-robot relationship.

As Artificial Intelligence permeates into both the social and creative spheres, research such as this becomes of great interest and importance, particularly as it deals with the intersection of the two. The blog you are reading may have been made by a computer program [Phactory, 2012]. The person you are instant messaging may be a chatbot [BBC, 2012]. The music you are listening to [Eacott, 2001] and the painting you have just seen [Brown, 2008] and the portrait that has just been drawn of you [Tresset and Leymarie, 2006] may have all been produced by a computer program or indeed, a robot. Time and again, humans and artificial agents continue to coexist in these endeavours and as many of these previously anthropro-dominated areas are breached by artificial beings, it does not seem so unlikely that the questions proposed by this research will yield positive results. This work opens up avenues in the study of human-robot relationships, providing the HRI community with a greater understanding of the place shared creativity, and specifically music, has in building and maintaining long-term engagement and social presence. Additionally, designers of Interactive Music System (IMS) gain insight into the role simulated social behaviours can play in user engagement.

1.2 Thesis Overview

We will now provide an overview of the thesis's structure and its contents. In Chapter 2 we survey the existing research fields relevant to the thesis statement presented in Section 1.1. As the quality of a human-robot relationship is what we expect to study and improve, we primarily cover social relationships between humans and non-humans, including robots and animals. Next, as it is with a strong foundation of engaging musical interaction we expect to improve human-robot relationships with, we detail the links between social relationships and musical activity. Also, how algorithms have been used to create formalised composition systems and how computers have allowed these systems to become interactive in both virtual and physically embodied forms. Finally, in search of existing models and methodologies that could be used to determine the quality of a human-robot relationship, we survey the main evaluation techniques of the HRI community and those used in the study of human-human relationships. These sections bring to light under-examined areas, gaps and challenges and inform the research questions this thesis proceeds to address. The research questions are explicitly set out in Section 2.4.

In order to address these, it was necessary to develop a method of evaluating the quality of human-robot relationships that may be used in the proceeding experimental HRI studies. Chapter 3 provides a model and methodological approach of automated behavioural metrics to be used for this purpose.

To further investigate our hypothesis that music may provide a solid foundation for human-robot relationships, in Chapter 4 we detail the use of an online version of the Network of Relationships Index Social Provisions Version (NRI-SPV) relationship survey to uncover the provisions of a relationship based in regular musical activity.

In Chapter 5, taking cues from the results of Chapter 4, we outline the development of *Mortimer*, a responsive drumming robot, designed, constructed and programmed by the author.

The next 3 chapters detail studies conducted with human pianists and *Mortimer*, progressively investigating whether additional social modalities improve the potential for a positive human-robot relationship centred around improvised music. Chapter 6 covers a single session study where *Mortimer* is presented as just an interactive instrument or as a simple social actor. Chapter 7 describes a long-term study where the addition of socially and musically triggered nonverbal behaviours is investigated. Finally, with another long-term study, Chapter 8 takes a look at extending the sessions beyond the laboratory with the addition of virtual presence.

Chapter 9 summarises the contributions of the thesis and discusses potential avenues for future work.

1.3 Associated Publications and Demonstrations

1.3.1 Publications

- L. McCallum, P. McOwan, Shut up and Play: Engagement and Social Presence Human-Robot Musical Interaction, in: Proc. Human and Robot Interactive Communication Symposium, Edinburgh, 2014, pp. 138143, *IEEE*. This paper covers research presented in Chapter 6 of this thesis.
- L. McCallum, P. McOwan, Face the Music and Glance: How Nonverbal Behaviour Improves Human-Robot Relationships Based in Music, in: Proc. Human-Robot Interaction Conf., Portland, OR, 2015, pp. 138143, *ACM New York*. This paper covers research presented in Chapter 7 of this thesis.
- Extending Human-Robot Relationships Based in Music with Virtual Presence, Robotics and Autonomous Systems Journal - Special Issue on Robotics and Creativity (Under Review), *Elsevier*. This article covers research presented in Chapter 8 of this thesis.

1.3.2 Demonstrations

- Robot Demonstration, Brighton Science Festival, 2014
- Talk and live demo, British Science Festival, Birmingham, 2014
- Robot Demonstration, Royal Institute Lates, London, 2014

- Research Documentary and Interview, Royal Institute, 2014
- Robot Demonstration, Festival of the Spoken Nerd Comedy Show, London, 2014
- Royal Institute Christmas Lectures, BBC, 2014
- Robot Demonstration, The Gadget Show, Channel Five, 2015
- Robot Demonstration, SMG Research Showcase, QMUL, 2015, Winner of Late Stage Research Prize

Chapter 2

Related Work

2.1 Social Relationships Between Humans and Non-humans

It is not an unreasonable assertion that social relationships are an integral part of both human's intelligence and their perception of the world on a day-to-day basis, also, that this is far from a modern affectation. Humans survive in groups, learn by tradition, trade and enjoy each others company [Dautenhahn, 1998]. Evidence points to as much as 99% of the last 2 million years of human existence having been spent in groups [Henderson, 1977]. There in still no consensus on the exact reasons for the evolutionary persistence of these behaviours, but some have even suggested that the need to solve social problems actually drove the development of human's intelligence.

While we are eating our picnic on this side of the uncanny valley, where robots are still distinguishable from humans, Dautenhahn notes that robots are **not** humans, and will be treated as such [Dautenhahn, 2007]. She astutely recognises that whilst observations have been recorded of human's social behaviour towards inanimate objects and animals, for example, the work of Reeves and Nass [Reeves and Nass, 1996], due to the innately social configuration of human intelligence this should not be surprising. As we are interested in examining if this predisposition may stretch to allow a positive and sustainable social relationship to develop between human and robot, we discuss below cases of perceived, projected and actual social relationships between humans and non-humans and the mechanisms through which this may occur.

2.1.1 Animals

Whilst there are clear suggestions that the companionship bond between human and animal is projected by the humans and their propensity to perceive relationships anthropomorphically, there is also evidence to suggest that animals can develop relationships with humans similar to same species social bonds. The study of this provides us with insight into the possibilities and limitations of social relationships between humans and non-humans.

Rajecki and Lee Rasmussen describe a study in which college student's perceptions of the mentality of a dog were compared to that of a fictional male child [Rajecki and Lee Rasmussen, 1995]. Participants were questioned regarding general mental processes and possibility for remorseful thoughts. Whilst the boy received higher ratings for more complex behaviours, overall, participants perceptions were judged as qualitatively similar. They suggest that although significantly subservient to humans with regards to intelligence, dogs are given near-equitable human social standing. This is supported by the study of caretakers and their dogs by Sanders [Sanders, 1993] that further strengthens the theory that nonhuman-human social bonds are made possible by humans innate desire to see things socially. A further example is the description of an interspecies play platform *Cat Cat Revolution* by Noz and An. They report the desire of humans to include pets into their gaming experience. Also, that they will actively attempt to interpret the interaction with the cat on their level in order to share a mutual meaning [Noz and An, 2011]. Again, the human attempts to balance an asymmetric relationship, this time by lowering themselves to a cats social skills, rather than project human traits onto it. All of the above strengthens our view that humans are capable of having social relationships with non-humans such as robots.

Endenburg and Hart describe an in depth study into motivations for the ownership of companionship animals in the Netherlands [Endenburg and Hart, 1994]. We feel these will be pertinent to our discussion as it may aid us in designing a robot in possession of characteristics desirable to humans. Participants relate social advantages such as tactile contact, attachment and dependency. In fact, 79% of reasons fall into this category. This is clear a portent of the ability, and desire, of humans to seek social relationships with pets. However, many participants were unable to accurately describe their relationship, again demonstrating the difficulties of studying a so commonly ill-defined concept. They further suggest that it is a different type of companionship people seek from pets, rather than a replacement for absent human relationships. This points that their is not only desire of human's to have relationships with non-humans, but that they may have different models for these relationships.

Communication, interpretation and comprehension between animal and human are all relevant to our study, as it informs us how a human would interpret, and what meanings they would ascribe, to the actions of a nonhuman, and because they illuminate how humans may naturally attempt to communicate with a nonhuman. For example, humans will regularly attempt to use gestures that they assume the dog can understand, as well as address their dog verbally for prolonged periods of time. Whilst dogs can be said to understand some utterances, it is definitely far from human level comprehension. For example, Pongracz et al. describe how owners believe their dogs interspecies vocabularies stretch up to 30 words [Pongrácz et al., 2001]. This is another case of humans treating animals socially. Twinned with findings that humans can accurately distinguish between different messages in dogs barks [Molnár et al., 2006], this shows humans will both try to communicate with, and understand, nonhuman companions.

Social relationships between humans and dogs are also not all in the mind of the human. Having evolved through artificial selection alongside humans, domestic dogs (*Canis Familiaris*) are social creatures themselves and have been bred to have companionship relationships with humans [Virnyia et al., 2004]. They are able to follow human gestures such as point cues and use visual cues to signal to humans [Miklósi et al., 2005].

In summary, animals, especially dogs, are capable of building social bonds with humans, which studies show they reciprocate. These bonds may provide benefits that are not just the same to a social relationship with a human, but different or better.

2.1.2 Robotic Pets

Drawing on the conclusions of Section 2.1.1, it would seem sensible that using the surrogate of a human-animal interaction may be an achievable first step towards human-robot relationships. The form of robotic pets may even cause them to be better received than android robots, as having an anthropomorphic form can hinder the positive reception of a robot by allowing users to presume currently unachievable levels of human-equitable social ability.

A study of forum posts from owners of Sony's robotic pet *AIBO* conducted by Friedman et al. discovered that many of its users (47%) referred to it with biological descriptors. However, this is not so surprising given its explicitly zoomorphic form [Friedman et al., 2003]. Less, but still a reasonable amount, go as far as to attribute animism or lifelike behaviours (14%). 59% mentioned they had some type of social rapport with the toy and within this 28% stated an emotional connection and 26% referred to companionship. Notably, the majority (75%) made some reference to the toy being an inanimate technological artefact, demonstrating they made all of the above comments despite explicit acknowledgements of the *AIBO* as a nonliving machine. Studies with the elderly and the robotic seal *Paro* in Giusti and Marti also find that acknowledgement of inanimacy does not serve as a barrier to pleasurable experience [Giusti and Marti, 2006]. Again, this provides support for our aim to build positive human-robot relationships.

The idea of using a zoomorphic animal with children in hospitals has been trialled by both Diaz et al. and Stiehl et al. [Diaz et al., 2011,Stiehl et al., 2009]. The latter introduce the *Huggable*, a robotic teddy bear for use in paediatric wards. Suggested uses included various health education applications, but also as a companion for hospitalised children. In a similar fashion to dog ownership, they suggest it may facilitate interactions with other children. Further, Diaz et al. conduct a study with the *Pleo* dinosaur platform, as well as the humanoid *Nao* robot [Diaz et al., 2011]. With regards to child-pet relationships, they recognise the interdependence of the master-pet dynamic is likely to produce affection, as well as care giving and teaching behaviours. This will be to the therapeutic benefit of the child. During studies, they reveal the large extent to which the appearance of a robot affects the expectations of functionality and attribution of social characteristics. This includes the distinction between between humanoid, mechanoid and zoomorphic and informs us that the form of a robot should be considered carefully when aiming to develop a relationship with a human. Whilst the response to *Pleo* is generally positive and comparable to the relationship with a non-robotic pet, they do also note that children desire *Pleo* to express more lifelike behaviours.

2.1.3 Sociable Robots

As robots migrated from conducting mainly manual and menial jobs in factories and warehouses to becoming more part of our everyday lives, it became necessary to embellish them with some level of social intelligence. There is within this, however, a broad range of motivations for such research. Zoll et al. note that the intended roles may include pet, butler, carer, playmate, coworker or even pseudo-child [Zoll et al., 2010]. This being said, often development has focussed on making robots socially aware and able to act appropriately in social situations in order to to cooperate better when carrying out functional tasks alongside humans. Although this approach is subtly distinct from one which seeks to develop robots as companions in the sense of kinship and close social bonds, the set of social faculties that need to be engineered for both tasks are closely aligned. As such, a survey of the important considerations in social robotics will greatly inform us in attempts to develop human-robot relationships.

Whilst any robot must have a basic level of functional competency to carry out its intended role, there are number design issues specifically raised by sociable robots. The most fundamental of these is real time performance, as if they are to interact with humans, they must be able to do so *in simuli* and in a natural and responsive manner [Breazeal, 2004]. A social robot should also be able to perceive human behaviour and activity, such as gesture, language and facial expression [Fong et al., 2003]. They should also be able to use these to determine the social relevance of any actions, affective states and context [Aylett et al., 2011].

In the field of affect sensitivity, Castellano et al. note that it will be necessary to go beyond primitive techniques such as camera based facial recognition of caricatured expressions, as these are often more subtle in real world situations [Castellano et al., 2010]. A smile, for instance, should not always be interpreted as a signifier of a content inner state and one action may not be the appropriate response in all scenarios. Continuing with the importance of context, Brezeal claims that to appropriately simulate human social perception, a robot must be able to sense not only others presence, but what they are doing and why they are doing it [Breazeal, 2004]. For example, the roboceptionist *Valerie* described by Gockley et al. uses a laser to detect not only proximity, but also attendance [Gockley et al., 2005]. This means she will only initiate interactions with those she deems to be focussing attention towards her, rather than all who are close.

It may also be useful for a robot to be able to identify a human's personality. Individual personality traits have been shown to affect the way a human interacts with a robot in several studies [Walters et al., 2005, Aly and Tapus, 2013, Dang and Tapus, 2014, Takayama and Pantofaru, 2009]. As such, if the robot can identify a personality type, they can adjust their behaviour appropriately and increase the chance of a positive social relationship developing.

Beyond perception, the ability to produce clear responses is also necessary. This allows the user to view any affective state and interact in a natural manner [Fong et al., 2003]. Pereira et al. subsume the notion of affective sensitivity into the wider concept of empathy, further encompassing the ability to take a perspective and to communicate a feeling of caring [Pereira et al., 2011]. They suggest that empathy is a necessary facet of a successfully social robot and find that users perceive an empathic robot as more of a friend than a non-empathic counterpart. Empathic behaviour has also been favoured in studies of virtually embodied relational agents, being cited as more integral to a system than user expressivity by Bickmore et al. [Bickmore and Schulman, 2007].

The appearance of a robotic companion can also have a profound effect on how itself and its behaviour are perceived. The phenomenon of the "uncanny valley" is where feelings of revulsion are heaped upon an anthropomorphic robot and is widely publicised, however, there are a few more subtle phenomena unearthed by studies of robot appearance. A deliverable report on the Foundations of Embodied Companion Surveys from multi-institution EU research project Living with Robots and Interactive Companions (LIREC) [Lirec, 2008] surveys academic output in this area and informs us that users would favour a robot that has human-like features. This being said, they do not want one that explicitly looks like a human. This distinction is a preference for humanoid over android robots.

Walters et al. found that humanoid robots were perceived as more intelligent than mechanoid robots, also that taller robots were perceived as more human-like [Walters et al., 2009]. Human-Robot Proxemics (HRP) is the study of the spatial relationship between the human and robot and can also be affected by a robot's appearance [Syrdal et al., 2008]. For example, how human-like a robot is perceived as being is inversely related to how close they wish it to get. Long-term studies into HRP by Walters et al. show that, after an initial settling period, preferences in human-robot distance tend to remain constant [Walters et al., 2011]. The results found in the above studies suggest that the appearance of our robot could have profound effects on preference and perception.

Personality is part of a 5 category model for human-robot companionship defined by Benyon et al. [Benyon and Mival, 2010] and one way a robot's personality can be embellished is through the use of apparent autobiographical knowledge. Dautenhahn notes that non-primate species, perhaps due to a lack of autobiographical memory, tend to take part in mainly anonymous interactions [Dautenhahn, 1998]. The implication is that autobiographical memory allows us to build a sense of self that we then use to relate to others and develop individually specific social bonds. This is succinctly summarised in the phrase "Ants don't tell stories" [1]. An attempt to include this in HRI is reported by Gockley et al., who collaborated with their Drama department to provide a progressing back-story for their robot. This was played out during real life interactions and via a website [Gockley et al., 2005].

Sarah the Facebot is another example of a robot attempting to develop their personality [Mavridis, 2010] and in this case it was by embellishing physical world interactions with a social web presence. She can personalise dialogue by querying the social network Facebook and gathering data about participants, and also publishes information about herself to the site. Moreover, being placed within this virtual social network, she can inquire about common friends, creating a sense of shared knowledge and context. Although results are positive, Bickmore et al. discuss the ethical implications of giving a fictitious back story to an agent [Bickmore et al., 2009]. They note that whilst this has the potential to erode trust, deception is common to all societies and necessary to many professions, and so may be acceptable. Further, they find rather that than renouncing simulated autobiographical stories as fraudulent, users became more engaged with agents that recounted such tales in the first person compared to those retelling about a third person.

Any social relationship is made up of a number of social interactions the research above shows that there are many things to consider when designing a robot to take part in social interactions with humans. Overall, we find that by increasing the social capabilities of a robot, humans become more engaged and this points to their importance in developing human-robot relationships.

2.1.4 Long Term Studies of Human-Robot Interaction

Studies in the field of HRI are often conducted as single sessions and seldom extend beyond this. With respect to human-robot relationships, this allows few

[1] [Dautenhahn, 1998]

satisfactory conclusions to be drawn as a social companion robot should not only be able to engage the user in a pleasing and natural interaction, but also be able to maintain this over an extended period [Campos and Paiva, 2011]. As will be detailed below, research shows even complex behaviours and intricately engineered robots become obvious and irritating for both children and adults over longer time frames [Fernaeus et al., 2010, Gockley et al., 2005]. This is referred to as the "10-h barrier"[2] by Tanaka et. al [Tanaka et al., 2007] and demonstrates a short-term novelty effect in HRI and inherent difficulties in creating long-term human-robot social relationships. Below we cover the long-term trials that have been conducted in a number of contexts.

Studies with Primitive Robots

Fernaeus et al. describe a long-term study with the robotic toy dinosaur *Pleo* [Fernaeus et al., 2010]. The toy was given to 6 families for periods of up to 10 months and the participants interviewed periodically. They report similar initial biological projections to the *Aibo* [Friedman et al., 2003] as reported in Section 2.1.1. Yet, despite the mechanical and sensory sophistication of the toy, it failed to maintain long-term interest. Issues arose involving a failure to live up to expected interactive behaviours across multiple modalities suggested by its price and design. Children favoured interactive over autonomous behaviours and found *Pleo's* apparent lack of awareness of context and reactivity to play resulted in a "mundane confusion" [3]. Whilst some of these disappointments can be attributed to the unrealistic expectations of children and their transitive relationship with toys in general, it highlights the challenges facing any attempt to build a robot which can even remain engaging over extended periods of time, let alone develop a strong social relationship.

Another study with a relatively simple robot was conducted by Sung et al. using the commercially available robotic vacuum cleaner *Roomba* [Sung et al., 2009]. They were motivated to uncover user's changing behaviour towards the system over time, especially after the novelty period. During a 6 month study with over 30 households, they detail a wide variety of positive responses, such as receiving unprompted updates and photographs of the Roomba in action. However, these reduced considerably after the introductory stages. They also note the problem of asking people to recall information about events which have become ingrained in their routines, seen initially after 2 months and then

[2] [Tanaka et al., 2007]
[3] [Fernaeus et al., 2010]

again after 6 months. They suggest the use of generative interventions to help participants report routine behaviours in long-term studies. This hints to the possible methodological challenges of conducting long-term HRI trials, which will be necessary when investigating human-robot relationships.

Studies in the Workplace

Researchers have also used the workplace as the testing ground for their work, seeing the co-worker as a potential role for the social robots of the future. Gockley et al. placed the *Valerie the Roboceptionist* into the foyer of their university for 9 months [Gockley et al., 2005]. She was able to sense when a passer-by had stopped and relayed a unfolding, serialised tale of her fictional life. They report many returning to interact with her over a 9 month period, however, her stories are relayed as monologues, with no personal link to the visitor, or ability to interact during. The negative effect of these limitations can be seen in participants regular departures before each 3 minute episode was completed. This suggests interactivity and personalisation are an important part in maintaining long-term engagement.

Mavridis et al. cite the aim of their FaceBots project as being to develop sustainable and meaningful long-term relationships between human and robots [Mavridis et al., 2011]. They attempt to achieve this by using facial recognition, natural language processing and by leveraging knowledge gained by situating the robot within participants online social networks. Also, they argue for the use of shared episodic memories between the human and robot to allow a relationship to grow over time. Whilst they report a six month study had taken place, the majority of evaluation relates to the facial recognition software and there is little discussion of how the robot was received or how its relationships developed. They describe a short study in which a non-embodied version of the robot held online chats with its virtual friends, however, there is little explanation of the outcome. It is possible the lack of a clear model and methodology for evaluating the quality of human-robot relationships limited their ability to study this aspect of the robot's interactions.

Mitsunaga et al. also executed a study into social interactions between humans and an anthropomorphic robot, one important addition being the use of haptic sensors [Mitsunaga et al., 2006]. Using the *Robovie-IV*, they allowed it to wander around their lengthy office corridor and identified four main categories of interactions, differing by who initiated it and whether a response was expected.

Similarly to the Facebots, it keeps a record of previous interactions to refer to in subsequent meetings. They reported that technical limitations lead to mistakes by the robot that resulted in confusion for workers. Further, this study also fell victim to the novelty effect and again demonstrates the issues that can arise when a social interaction is conducted even slightly incorrectly.

Whilst in the two previous studies the robot's main role was to socialise, others attempt to develop relationships around a functional task. For example, *CERO* was used to execute transport tasks for a motion-impaired employee, although without any explicitly social behaviours [Huttenrauch and Severinson Eklundh, 2002]. Another was conducted by Lee et al. with the SnackBot delivery robot [Lee et al., 2012]. Again, the robot used information from previous interactions to build the relationship over time. One important difference in this research is that the robot arrived at specific times during the week to carry out a well-defined task, rather than wandering around, as with the Robovie-IV, or having no direct purpose, as with *Pleo*.

They describe two 2 month studies in which social dynamics grew between users and the robot, finding that rapport personalised to the users increased both cooperation with the robot and anthropomorphic descriptions of the robot during interviews. Although people did have positive reactions, building anticipation of the robot's arrival into their routines, others felt the social interactions were sometimes unwarranted and unhelpful. If they were busy the robot had no way of detecting this and carried on regardless. Awkwardness similar to the Robovie-IV study was also noted when the robot said things that did not make sense or failed to end an interaction at the appropriate time. It is worth noting these effects were noticed even though the experiments were partially controlled using a Wizard of Oz (WoZ) technique, highlighting how far current technology falls short of the necessary mark.

Studies in Education

Kasap and Magenant-Thalman describe a study examining a robotic tutor-human student relationship over 4 sessions in a 2 week period [Kasap and Magnenat-Thalmann, 2011]. They use a highly anthropomorphic robotic bust named *Eva* and found it was able to maintain its social presence across the 4 sessions. They also find that the inclusion of memory into the dialogue had a significant effect on maintaining a users engagement over time and find a link between matching personalities for user and tutor and task motivation. Whilst this study does present some successful techniques for maintaining user engagement over multiple sessions, it is arguable whether 4 sessions over 2 weeks with 1 hour total interaction time constitutes a long-term relationship.

Social presence is also suggested as a factor in motivating humans to maintain relationships with social robots by Leite et al. [Leite et al., 2009]. Again using the student-tutor scenario, they introduced the Phillips *iCat* into a children's chess club and saw that the social presence decreased over the 5 weekly sessions. From watching the children interact with each other, they suggest that adding memory, as was done in the *Eva* study, may improve results.

Working with children who are younger, and so less developed, has shown more positive results. In a 5 month study with toddlers between 18 and 24 months old, Tanaka et al. demonstrated increasing quality of interaction for a sustained period of the trial [Tanaka et al., 2007]. Although there was some human supervision of the robot, they suggest this was minimal and that the *Qiro* robot was almost ready to develop fully autonomous relationships with young children.

Sociable robots also provide promising results when used alongside children with conditions that impair their social interaction skills, such as autism. For example, the robot KASPAR was shown to improve collaborative play between children with autism following several interactive sessions [Wainer et al., 2014]. This study not only demonstrated engaging interactions between the children and the robot during the trial but also sustained therapeutic effects afterwards. Furter, non-directive play sessions with a robotic pet have also been shown to have a positive effect on autistic children's ability to play and reason [Francois et al., 2009].

2.2 Music as an Approach to Human-Robot Relationships

Section 2.1 demonstrated that humans are capable of taking part in social interactions with non-humans, sometimes even developing relationships. However, from the studies discussed in Section 2.1.4, we find that when the counterpart is a machine, this is often limited by the choice of interaction domain and design of morphology and behaviour in current sociable robots.

As part of the thesis statement in Section 1.1, it is our hypothesis that open-ended musical activity can provide the necessary engagement to begin and develop a human-robot relationship. In support of this we cover the role music can play in the development of human-human social relationships. Proceedingly, we survey the necessary technical fields in building a robot capable of improvising in realtime with an acoustic instrument.

2.2.1 Social Relationships and Music

Any time music is played as part of an ensemble, you are guaranteed to have at least two people, in the majority of cases co-located, simultaneously focussing their attention towards the same task and cooperating towards a joint goal. Without making much of a leap, we can already suggest that this fulfils the condition of shared interest to foster social bonding first set out by Robert Weiss [Weiss, 1974]. This type of activity, often routine, regardless of the musical context, will clearly develop some type of familiarity. However, whether this alone is sufficient to develop a relationship that can be characterised as a social bond is questionable. We will see how, and to what extent, the practice of participative music making can fulfil this further.

Kokotsaki and Hallam ran a thorough qualitative evaluation of the perceived benefits of participative music making by Higher Education students [Kokotsaki and Hallam, 2007]. Whilst many of the advantages are practical, such as development of musical skills and increased ability to express emotion through music, there are a plethora of social benefits extolled by the participants. Some points raised can be broadly grouped into feelings of inclusion, such as the perception of themselves providing an important contribution to the group, feelings of pride in group success and a sense of belonging. Others refer to the development of transferrable social skills like compromise, mutual support and encouragement. The alignment of these benefits with the provisions of a social bond along the axes of positive feedback, guidance and being valued begins to demonstrate how music making can be a steady platform for developing strong social relationships.

Looking closer into this, we see that whilst music can be a solitary pursuit, it is often communally enacted. By extrapolating from contemporary and aboriginal musical practice, Dissanayake suggests music may have even developed in early humans through social activity as a communal ritual [Dissanayake, 2008]. If it is true that humans have an evolved propensity for musical behaviour, and that this was in some way adapted from social behaviour, this would demonstrate not only a strong and long standing link between musical and social activities, but also an almost causal relationship. Dissanayake does correctly, however, note the problematic nature of characterising contemporary aboriginal peoples as ancient hunter-gathers.

Further, similarity in music preferences has also been shown to be positively related to formation of social relationships and friendship [Selfhout et al., 2009] and whilst music preference is strongly linked to personality traits, Boer et al. suggest that it is matched values rather than matched personalities which form the basis for this social attraction [Boer et al., 2011]. The theory of interpersonal attraction posits value agreement as mutually advantageous, whereas disagreement is a threat to one's perception of the world and thus is disfavoured. With this, their evidence demonstrated that those with similar musical preferences were more likely to share personal values, and so were more likely to form social bonds. They also found their results to hold across Western and non-Western cultures.

Studies have shown that people, especially adolescents, use music as mood management [Hargreaves and North, 1999], therefore, if the robot can provide a positive effect on affect, it may help foster a long-term relationship between human and robot. Further research suggests humans also use music to develop and maintain their self-identity [Hargreaves and North, 1999] and there is also evidence to suggest humans are more open to interacting with a robot that displays compelling personality traits [Walters et al., 2009]. Therefore, music may aid the development of human-robot relationships as there is the potential to display a unique personality and identity.

The research above demonstrates that shared musical experiences can be a strong platform for developing social relationships. This includes actual ensemble performance or simply aligned musical interests.

2.2.2 Interactive Music Systems

Up until the midpoint of the preceding century, performance and composition of music had been almost exclusively the pursuit of humans. Increasing power and availability of computers, not just in research laboratories but also in many homes, rucksacks, and even mobile devices has brought forth the use of artificial agents for these ends into the realms of possibility. Although since then, research in developing artificial systems which can perform, compose and understand music has progressed at a steady rate, these advances still leave us a fair distance from musically autonomous agents. The intersection of these three areas provides the necessary competencies for any IMS, to be defined as one able to interpret musical input and generate musical output as realtime response [Winkler, 2000]. HRI trials that have treated creative tasks as in a crass and simplified manner have failed to engage participants [Tanaka et al., 2007, Kose-Bagci et al., 2007], so we acknowledge that if we are to move forward in the task at hand we must properly appreciate the subtle and complex nature of a musical interaction. As such, if a musical interaction is to provide the main engagement for a human-robot relationship, an understanding of the achievements, limitations and challenges of IMS research is crucial.

We would like to acknowledge the panoply of research fields, including machine listening, algorithmic composition, interactive music systems and computational creativity, to name a few, whose actions and words have potentially great relevance to the project being undertaken, each with their own methodologies, motives, arguments and controversies. However, as previously stated, the aims of this project, and so the grounds upon which success will be judged, lie not in the actual output of the humans and machine, but to the extent engagement and social presence can be developed and maintained. That is not to say our research will not draw heavily from, nor fail to contribute to, these areas, just that we cannot hope to recount and address all issues raised by each's endeavours.

Algorithmic Composition

Pearce et al. note that the majority of projects that use computers to implement formalised systems of composition can be categorised somewhere on a four way spectrum of algorithmic composition consisting of compositional tools, modelling of musical styles and modelling of musical cognition [Pearce et al., 2002]. From these, it is thought that the first two present the most relevance to this research and so will receive the most detailed coverage. Formalised systems for generating music predate the use of computers by almost a millennium. For example, in perhaps the earliest example of data sonification, Guido d'Arezzo developed rule based system for turning lines of text into melodic scores in 1026 [Roads, 1996]. Moving forward, Western Tonal Music is in itself highly formalised and lends itself generously to composition by mathematical rule based systems. Mozart is said to have generated a series of cards for a dice game for his *Musikalisches Wurfspiel*, capable of composing a great number of new pieces by stochastic rearrangement of phrases [Nierhaus, 2009]. Moving further forwards to the start of the twentieth century, Arnold Schoenberg introduced a serial system to control the pitches of his compositions. This was later furthered by von Webern to control even more aspects of his pieces, such as rhythm and dynamics [Edwards, 2011]. Other examples of rule-based generative systems outside of the classical canon include John Cage's experiments with Chinese mysticism and the I-Ching, notably in his collage *Williams Mix* or *Music of Changes*. The coin tossing involved in the composition of the latter took reportedly 9 months [Muscutt, 2007]. Further indeterminacy in performance and music can be noted in the works of LaMonte Young [Lee Martin, 2002] and Yoko Ono [Ono, 1964]. All of these show that even without a computer, composition can thought of it terms of generative procedures and musical pieces in terms abstract instructions. Both these will be necessary to build a robot capable of composing music responsively, as opposed to one that simply replays precomposed pieces from a score.

Including a large number of features into the algorithmic process increased the calculative load and the use of these systems to compose by hand became rapidly time consuming; impossible, even. When faced by this situation in the present day, it is common to farm out the computation to a computer. However, it was not until machines were powerful and available enough that was this attempted. The *Illiac Suite* by Lejaren Hiller is often cited as the first piece of music composed by a computer [Pearce et al., 2002, Roads, 1996, Edwards, 2011, Collins, 2010]. This collection of four pieces for string quartet was produced by the Illiac Computer at the University of Illinois by a program written in binary machine code. It was closely followed by *Push-Button Bertha*, a song written by computer for popular audience [Holmes, 1985]. Another innovator of computational composition and sequencing was Raymond Scott, whose experiments in the 1950s lead to numerous computationally generated advertising jingles, introducing electronic music into the public realm well before its widespread use in rock and pop [Moog, 2012].

As we will come to in Chapter 5, in terms of proliferation it may be desirable to approach popular music as a field for building human-robot relationships. However, despite these early commercial forays, it was experimental electronic music which took up algorithmic composition with most fervour. Notably, Iannis Xenakis relied heavily on computationally executed algorithms and pseudorandom stochastic processes in the 1960s. His most prevalent system was the Stochastic Music Program [Xenakis, 1992]. As alluded to earlier, Western orchestral music and its harmonic and compositional structures makes the use of algorithms for all or part the process not too great a leap in practise. Probably the most famous proponent in this field is David Cope, whose EMI system can be trained on a corpus of music in a particular style and produce similar compositions [Muscutt, 2007].

A more recent example of algorithmic composition as a novel distribution tool is described by John Eacott [Eacott, 2001]. The CD ROM *MORPHEUS≫ emergent music (contents may vary)* from 2001 provides 16 tracks in algorithm, rather than concrete, form. Thus, each time the CD is listened to, the experience will be a novel one. He answers the criticism of algorithmic music's transitive nature meaning no one can ever really get to know a song with a 'humming test', which he provides anecdotal evidence of the CD passing. He also notes that what he refers to as fluid dance music bridges the gap between recorded and live music in its variance, and in doing so inherits both its benefits and deficiencies.

Interactive Music Systems

By definition, an IMS must be able to respond in realtime. This constraint was also previously identified for sociable robots. This was beyond majority of systems described in Section 2.2.2 as it computationally feasible until the 1980s. Breaking this barrier was also aided by the introduction of formalised sequencing protocols such as Musical Instrument Digital Interface (MIDI). These progressions allowed the introduction of algorithms into musical instruments and the twinning of algorithmic composition systems with live data from other performers, be it from the audio of an accompanying musician or even the movements of a dancer.

Responsiveness and empathetic behaviours greatly increase engagement in HRI [Pereira et al., 2011,Bickmore and Schulman, 2007]. Consequently, building a robot that can respond to a participant's playing in realtime will be complementary to our goals of developing positive human-robot relationships. Projects involving interactivity will be further discussed below with a focus on systems designed for improvisation.

Robert Rowe declares that any system able to take part in an improvisation must, necessarily, be aware of external context and be able to response appropriately. This is in contrast to playing back musical segments randomly [Rowe, 1996]. Whilst Rowe is happy with a 3 step description of an IMS, distinguishing the *sensing* (data collection), the *processing* (interpretation) and the *response* (performance) stages, Todd Winkler extends the definition to 5 steps that we feel captures the whole process more completely.

> 1. Human input, instruments - Human activity is translated into digital information and sent to the computer.
> 2. Computer Listening, performance analysis - The computer receives the human input and analyzes the performance information for timing, pitch, dynamics, or other music characteristics
> 3. Interpretation - The software interprets the computer listener information, generating data that will influence the composition
> 4. Computer composition - Composition processes, responsible for all aspects of the computer generated music, are based on the results of the computer's interpretation of the performance
> 5. Sound generation and output, performance - The computer plays the music using sounds created internally, or by sending musical information to devices that generate sound [4]

This continuous feedback loop represents a realtime dialogue between human and machine which, whilst somewhat similar to a cycle of listening, adjusting and re-rendering an algorithm, offers a whole new cascade of musical possibilities, performances and experiences that are especially relevant to improvisational music. Robert Rowe also introduces the player-instrument paradigm for IMS. A system moving towards the instrument paradigm will be more like an extension of an instrument, whereas a system characterised as a player will be more analogous to a distinct musical performer. Clearly, a robot constructed to begin and develop a relationship with a human would tend towards the autonomy of the player paradigm.

Winkler remarks that some of the first interactive pieces predated the computer, using either feedback loops, such as Gordon Mumma's *Hornpipe* or voltage controlled synthesisers [Winkler, 2000]. For example, Morton Subotnick's *Touch* mapped the envelope of his voice to control synthesiser parameters entirely in the analog domain. Both Rowe and Winkler note the importance of

[4] [Winkler, 2000]

increased computing power and the introduction of the MIDI protocol. This allowed personal computers to take in expressive musical input in realtime, albeit at the cost of detailed timbral information.

One early pioneer of note is trombonist George Lewis, long a champion of improvisation over composition, who began developing his *Voyager* system in the 1980s. All versions were written by himself in various manifestations of Charles Moores' Forth language in an era before higher-level music specific environments such as SuperCollider and Max/MSP [Lewis, 2000]. *Voyager* was influenced Lewis's involvement in the Association for the Advancement of Creative Musicians, a group who exploited multi-layered improvisation to build rich timbres. Multiple agents exist as a virtual ensemble within the system and are created or removed aperiodically. These agent's general behaviours are then altered with regards to their choice of scale, timbre and tempo.

Voyager can either take input as a MIDI stream, or after a process of pitch detection represent real instrumentation, however, the agents can choose to follow, oppose or completely disregard these cues. If they choose to follow, an agents response is based on a general average of features such as timbre, tempo and note density. Once more Lewis shows a preference for clash, discord and ambiguity by and large avoided by most developers of IMSs who assign an almost totemic virtue to exact music transcription and traditionally appropriate response. An interesting emergent behaviour of this feature is that *Voyager* can start playing independently of external user input, allowing it an autonomy not usually afforded to IMSs. Further, rather than attempting to elicit a response based on inbuilt compositional and harmonic preconceptions, Lewiss system hopes that the improvisers actual emotional state will be reflected by the computer. This type of affective connection may be crucial to developing longer term bonds, although such unstructured improvisation may not be appropriate for our system.

In his doctoral thesis, Nick Collins describes 5 systems for computer music performance, 2 of which are not entirely autonomous [Collins, 2006]. Much of Collins's technical work on algorithms for autonomous computer musicians and thoughtful consideration of the limitations and advantages of placing a computer in a live ensemble is highly relevant to the intentions of this thesis. For example, *Free Improvisation Simulation* is a quintet of four artificial agents and one human guitarist developed explicitly for free improvisation. To inform response the system keeps some local memory of user's input in a motif database, however, in assessment of his experiences, the human player Ian Cross describes the system's limited ability to track larger scale structures. He also notes that all free improvisations sometimes work and sometimes do not. This demonstrates that considering the aims of this project, it may be worth sacrificing the freedom of expressivity allowed by such open-ended contexts to avoid an ungratifying interaction. In addition to this, *Sat at Sitar* is an improvisational performance for computer and human sitar player, involving realtime pitch and event tracking algorithms.

Francois Pachet proposed *The Continuator* as a system that would connect the often distinct categories of IMSs and music imitation systems [Pachet, 2003]. He suggests that previously, IMSs were unable to create stylistically consistent music, in that the output was heavily linked to the human's input to a much greater extent than to any consistent, recognised musical style. Conversely, he cites that music imitation systems, whilst able to generate music well in a particular style, rarely allow for interaction, thus limiting their use as an instrument. The presented system takes MIDI input, parses it into phrases used to form a model of the player's input. An extension of a Markov model, this is then used to generate *continuations* for the player. Pachet lauds this approach as being able to gain a feel for patterns played to it, but also to avoid typical problems associated with using Markov models to generate music with regards to longer term form. Here, this is neatly side-stepped as the user is in control of the extended form of the improvisation, while the system fills in gaps and provides responses locally.

The Continuator holds many interesting and potentially desirable properties with regards to the type of interaction we wish to illicit. Primarily, the introduction of learning into the system allows for an extended, and widely varied, personalisation process. Mentioned above, studies by Bickmore et al. suggest that empathic behaviour should be favoured over user-expressivity, and so, responses generated from a model learnt from user input can be seen to fulfil this requirement [Bickmore and Schulman, 2007]. In addition, its empty slate starting point enforces no particular style onto the user. Obviously, the input instrument itself affords a certain way of playing and the combination of input and output timbre also afford certain styles, however, these are not extant in the compositional process itself. This allows the system to mould to each new user, an approach which can avoid problems with lack of personal engagement perceived in many attempts at building human-robot bonds.

Mimi is a improvisation program working from MIDI input [François et al., 2011]. It is conceptually analogous to *The Continuator* in that it generates content based on a model of the user's input. A distinguishing feature is its visual interface, displaying the system's intended output from 10 seconds into the future. It also aids the human's memory of the structure of the improvisation on a wider scale with a visualisation of both user and computer's previous playing. Not only is this functionality usually unavailable in other IMSs, but also in human-human improvisation. Whilst they may have planned ahead, a human is not able to give such specific information about future contributions and in this way, *Mimi* brings qualities to an improvisation that are, arguably, impossible for a human to emulate.

Mechanical Instruments and Robotic Musicianship

So far in this section we have covered ways in which the composition of music can be formalised into algorithms and carried out by computational systems. Further, how these can take input in realtime from a human musician to inform their output. Finally, we cover the placement of these into physically embodied robots able to play acoustic instruments expressively. It is the combination of these three things that will provide us with a platform to build a robot capable of developing a positive and sustainable social relationship with a human through improvised musical activity.

Mechanical instruments and Music Performance Robots (MPR) are not a recent occurrence. Indeed, robots, or automata, have existed in the minds of engineers and the workshops of inventors for many millennia before Czech theatre coined the term in the early 20th Century. Generally built to imitate nature, early automata were impressive feats of design, manufacture and mechanics. Some of the earliest recorded came from arguably the two greatest technologists of antiquity, the Greeks and the Egyptians. The latter are said to have developed a human-like horn player powered by a clock as early as 1500BC [Culbertson, 1963]. This combination musical performance and physical automation is seen from the very first experiments and is repeated throughout the development of early automata. In these stages, they were seen as machines to entertain, rather than the tools of automated industry projected to replace humans in the dystopian futures that would be later imagined in Karel Capeks *R.U.R (Rossums Universal Robots)* and Fritz Langs *Metropolis*. Later examples include Jacques de Vaucansons hugely popular *The Tambourine Player* automaton in the eighteenth century [Hugill, 2008] and Friedrich Kaufmann's *Trumpet*

Automaton from 1810 [Ord-Hume, 1973]. The use of musical performance to display the aptitude of a robot continues to the present day, with Quadcopters used to perform the James Bond theme tune to great internet aplomb [of Pennsylvania, 2014].

Exterior to automata, Fabio suggests the earliest example of an instrument playing itself is the *hydraulis* pipe organ made by inventor Ctesibius of Alexandria [Fabio, 2007]. Beyond this, both Fabio and Kapur place weight on the invention of the player piano in the history and development of Robotic Music [Fabio, 2007,Kapur, 2005]. A precursor to the phonograph, this was the first widely popular platform for performance without human musicians in person. George Antheil is often credited as a pioneer of the creative use of this device with his composition *Ballet Mechanique*. It was scored for three xylophones, four bass drums, two pianists, a tam-tam, a set of electric bells, a siren, and three propellors, as well as 16 synchronised player pianos. Unfortunately the facility to synchronise such a mass of pianolas was not achieved within his lifetime, so he was never able to hear it in its full glory. The piece was resurrected to great acclaim in 1999 by Paul Lehrman. Previously replaying percussion through samplers, this version was later fully automated with the help of The League of Electronic Musical Urban Robots (LEMUR)'s Eric Singer, in the National Gallery of Art in Washington, DC in 2006 [Lehrman and Singer, 2006].

LEMUR have developed and displayed a panoply of original and innovative musical instruments over the past decade, the most notable being its debutant, the *Guitarbot*. This instrument consists of four guitar strings, each with its own automated, sliding pick allowing for fast and continuous pitch movement [Singer and Feddersen, 2004]. Other robots in its arsenal include *TibetBot*, based around solenoid driven tibetan singing bowls, the rattling 10 armed *ForestBot* and the percussive collection of *ModBots*. Singer also worked with popular musician Pat Metheny to produce a fully automated backing band, christened *Orchestrion*.

A number of large ensembles of synchronised musical instruments have also appeared in recent years. Felix Thorne's *Felix's Machines*, originally presented as an installation in London, has toured the world, been used in concert by successful electronic music artist Plaid and has released a full length album of original compositions on Mute Records. Thorne uses his ensemble to compose and perform acoustic music, influenced by digital electronica artists [Thorne, 2012]. Another platform where an automated band is presented as the performer is Ghent University's Man and Machine Orchestra. Maes et al. provide a full and detailed account of the vast array of automated instruments on offer, consisting of those analogous to existing acoustic instrumentation, as well as newly designed noise generators [Maes et al., 2011]. The ensemble is motivated to provide a reliable and durable platform for various composers to write and perform music for automated instruments. Another is the Machine Orchestra, a student built ensemble of custom networked musical robots at CalArts. This is partly a teaching tool and partly a platform for composition and performance [Kapur et al., 2011].

Currently the only institution to boast a research group solely devoted to development of musical robots (the Robotic Musicianship group), the Georgia Institute of Technology leads the way in the academic study of human robot musical interaction. Their first, and most widely recognised, robot is *Haile*. A percussive robot, *Haile* is equipped with a real time beat tracking module and two beater-arms capable of expressively collaborating with a human player on a Native American Pow Wow drum [Weinberg and Driscoll, 2006]. Unlike many of the machines previously mentioned that were designed exclusively for autonomous performance, *Haile* can be classed as a physically embodied IMS. Weinberg and Driscoll see the limitations of many IMSs is that they do not provide ample visual feedback to actions and loose the rich sonic palette of acoustic instrumentation. As such, they were motivated to produce a robot capable combating these deficiencies. *Haile* is provided with a recognisably anthropomorphic wooden exterior, yet the majority of his mechanics are still visible, displaying how he works to the user. Further, LEDs are placed about the body to provide additional visual cues [Weinberg and Driscoll, 2007]. Rather than providing a platform for composers to write new music, they aim to stimulate 'inspiring human machine collaboration' [5].

Their next charge is *Shimon*, a robotic marimba player. Weinberg et al. describe *Shimon* as a 'Social Robotic Musician' [6], validating this with the description of a social module to provide visual cues for human participants via a screen based animated head [Weinberg et al., 2009]. Later models include a physical robotic head and neck. The primary function of this system is to provide visual cues, such as head bobbing, in time to the music and allow the robot to focus on different musicians, much as a human would during an improvisation. Similarly, Theofilis et al. describe a study using music as a communicative

[5] [Weinberg and Driscoll, 2006]
[6] [Weinberg et al., 2009]

gesture as the iCub robot is used to play drums with a human [Theofilis et al., 2013]. However, their evaluation of the social interaction is purely anecdotal so it is clear more thorough research is required in this area.

Finally, there is *Travis*, described as a "robotic musical companion". This robot consists of a smartphone speaker dock with the ability to move in time to your music collection, as well as find songs in your collection based on rhythmic patterns played to it [Hoffman and Vanunu, 2013]. They report positive effects of dancing on song liking, impression of agency and impression of similarity. The main interaction revolves around the shared focus on a person's musical collection of prerecorded tracks, rather than music making itself, so it is unclear whether these effects would be transferrable.

The Georgia Tech robots are aware of musical context and are able to react to this in realtime, fulfilling in some part the brief we have set ourselves in developing human-robot relationships. For example, *Haile* may keep in time and respond compositionally to the user's input using the measure of note density and *Shimon* will models 'gestures' of the human piano playing [Hoffman and Weinberg, 2010]. They also provide a social presence on stage, both to the benefit of the other players and audience. That being said, whilst *Shimon* may provide extra-musical multimodal cues, these are limited to during the musical interaction itself. A more holistic system may endeavour to generate some semblance of personality outside of the musical sessions.

2.3 Evaluating Human-Robot Relationships

In order for us to judge the existence and quality of a human-robot relationship developed through musical interaction, we will need to decide upon a methodological approach. In this section, we examine the plethora of ways researchers evaluate HRI and human-human relationships in search of an existing approach to adopt.

In all cases similar issues are raised, such as the balance of ecological validity with experimental rigour or the choice between expensive, obtrusive yet rich observation and cheap, quick but potentially questionable self report. Although most are equally apparent in both, there are some unique issues raised solely when dealing with the interaction of human and machine.

2.3.1 Evaluating Human-Robot Interaction

Dautenhahn notes the relative youth of the HRI field means that any standardised, shared methodology is yet to arise [Dautenhahn, 2007]. This had lead to HRI methodology often being borrowed and adapted from human-computer interaction, psychology and the social sciences [Bethel et al., 2007]. Indeed, it is the main tools of these disciplines such as self assessment, behavioural observation, psychophysiological measures, interviews and task performance metrics that have been adopted by HRI [Bethel and Murphy, 2010]. Therefore, although we have maintained throughout that the relationships that develop between human and robot are highly unlikely to be directly analogous to human-human social relationships, this does not mean methodologies and evaluative procedures from these fields may not guide us. For example, Ganster et al. recommend using existing standardised methods from psychology, noting that ad hoc questionnaires developed by robotic engineers can lead to naive assumptions and rarely allow for research to be compared [Ganster et al., 2010].

Several survey methods are currently in use in HRI trials. For example, Syrdal et al. describe a live trial using the Negative Attitudes Towards Robots Scale (NARS) [Syrdal et al., 2009], a questionnaire developed to gauge participant's anxieties towards robots. However, Suzuki and Umemuro note the limitation of NARS is it only tracks negative response [Suzuki and Umemuro, 2012]. To counter this they primarily use a group interview to gather potential attitudes towards robots, then use this data to compose questionnaires. Veenstra and Evers describe an online survey specifically for working with children [Veenstra and Evers, 2011]. The KidSAR tool is short in length and combines visual stimulus with simple text to interrogates children with regards to 10 different perceptions of robots.

Bartneck et al. present the Godspeed Questionnaire, 5 validated questionnaires to measure animacy, anthropomorphism, likeability, perceived intelligence and perceived safety of a robot [Bartneck et al., 2008]. Whilst these measures would not necessarily aid us in determining the existence of a social relationship, perhaps with the exception of likeability, they may help us explain the causes of any relationship should it be found. Moreover, they are to be commended for taking the approach of building properly validated questionnaires that provide high consistency in results.

In contention to these approaches, Bethel et al. outline the issues involved with self report in HRI trials, primarily that participants may answer what they believe they are expected to answer, rather than what they actually think [Bethel et al., 2007]. They are also critical of observational behavioural studies, noting the *Hawthorne Effect* when participants know they are being observed change their behaviour. To counter this they suggest the combination of one of the above with psychophysical information, as participants are less able to consciously manipulate these measures. One notable use of physiological measures is the work of Wada and Shibata, who repeatedly took urine samples from participants in a study with the elderly and *Paro* to evaluate possible stress reduction during therapy with the robotic seal [Wada and Shibata, 2006]. However, Bartneck et al. are critical of physiological measurements, noting that arousal from joy cannot be distinguished from arousal from anger using just this information [Bartneck et al., 2008].

Its seems this approach is limited both by the lack of related research directly in HRI and a lack of consensus as to which psychophysical signals are appropriate for use in HRI trials. One avenue beyond HRI that could prove fruitful is the physiological measurements of emotions in the field of Affective Computing. These measurements can include heart-rate, breathing rates and galvanic skin response [Picard et al., 2001] and have potential to be used when studying human's affective response to robots. Although sometimes considered invasive, recent developments have allowed the former two metrics to be gained from just the accelerometers on a user's smartphone [Hernandez et al., 2015]. Although promising, any attempts to classify affect will only ever be as strong as the emotional model used and the mapping from measured phenomena to that model. This is an as yet unsolved problem is equally applicable to psychophysical measurements of emotion.

There are also some issues posed solely by long-term research. Walters et al. report how 2 participants left their long-term study as they were bored with the repetitive procedures. This demonstrates the tension between maintaining rigorous experimental consistency and allowing for natural interactions [Walters et al., 2011]. Long-term studies may rely on qualitative data, such Fernaeus et al.'s research with the previously mentioned *Pleo* platform [Fernaeus et al., 2010]. The toy was given to 6 participant families, each were given a camera and encouraged to self report about their experiences, as well as film their play sessions. Each family also took part in a semi-structured interview at least once during the period. Other studies, such as the one conducted by Gockley et al., relied more heavily on quantitative data, such as interaction frequency and repeat interactions [Gockley et al., 2005].

Researchers have also developed techniques to evaluating HRI that do not include live studies with autonomous robots. This means that HRI trials can be done on robotic behaviours that have not yet been developed. Approaches to this are WoZ and Theatrical Robot (TR) [Dautenhahn, 2014]. In WoZ studies, unbeknown to participants, the robot is controlled by a concealed human and in TR studies, a human dresses like a robot and acts out pre-scripted behaviours. Video-based trials Video based Human-Robot Interaction (VHRI), where people are shown films of humans interacting with robots, provide the advantages of WoZ along with reduced running costs and the potential for remote distribution.

We have seen there are many approaches to both what to measure and how to measure it. The choices made by any researcher will ultimately be related to the particular slice of HRI they wish to investigate and the resources at their disposal. For example, Breazeal tells us that although some components of human-robot social interactions can be objectively evaluated, others are ultimately subjective [Breazeal, 2004]. In evidence she questions how useful an empirical metric may be for certain aspects of sociability, such as empathy. Unfortunately for us, most of the methods presented are of these objective aspects. Further, HRI shows a disappointing reliance on self report which is problematic when attempting to uncover the phenomena that may represent a positive human-robot relationship.

2.3.2 Evaluating Relationships

Due to a bulk of research into human-human interaction emanating from psychology-based disciplines, especially that relating to friendship and attachment, the two main approaches taken by previous researchers to evaluate social relationships are self report and observation. Self report has mainly been the approach taken to evaluating adult relationships and observational methodologies used for infants. Each have their own positive and negative attributes and the balance between them is described as the "trade-off problem"[7] by William Ickes. He warns of favouring one person's perspective and suggests getting as many perspectives from those within and outside of the relationship to gain the

[7] [Ickes, 2000]

fullest picture [Ickes, 2000]. Another trade-off is often between the levels of obstrusiveness of your method. Inherently, video observation in a natural setting will be less obtrusive than a gaze-tracking headset worn by each participant in a lab while researchers in white coats closely inspect their behaviour.

Those critical of self report often cite an over-reliance on subjective self-assessment from the participant. This can result in egocentric or ego defensive reactions attempting that either promote the participants actions or downplay others. Brennan et al. suggest that these response biases will be especially heightened when disclosing self insights with regards to fear and defence [Brennan et al., 1998] and Ickes notes that self report may be an issue if your subject has reason to lie. This may be the case with us as questions regarding feelings towards something consciously known as inanimate may provoke embarrassment in participants [Ickes, 2000]. For example, Reeves and Nass report people vehemently denying behaviours they have been observed doing after the event during their studies into social responses to technology [Reeves and Nass, 1996].

However, Furman and Buhrmester [Furman and Buhrmester, 1985] claim that self report can be vital when examining a social relationship. For instance, an insider's viewpoint on a friendship can put behaviours within a broader context through knowledge of past events and expectations of future interactions. To counter these issues, Bagwell [Bagwell et al., 2005] has attempted to combine both in a study of friendship quality, proposing that self report is necessary for identifying behaviours that are a rare occurrence, such as conflict, whilst observation can reveal behavioural manifestations of a friendship self report cannot. As they are less anonymous, face-to-face interviews are often preferred to phone interviews when sensitive questions, such as those regarding relationship satisfaction, are being asked. This can help in gaining detailed and personal information about a relationship [Ickes, 2000].

Several surveys exist to measure social relationships and whilst many focus on romantic relationships, others are available to test friendship or can be easily reapplied to different contexts. Previously detailed and used in Chapter 4, Furman and Buhrmester's Network of Relationships Index (NRI) is a general purpose model for measuring aspects of different relationships within a person's life [Furman and Buhrmester, 1985]. To some part based on Weiss' provisions of a social relationship [Weiss, 1974], the questionnaire also seeks to reveal non-positive attributes along the axes of relative power and conflict, with ratings done on a 5 point Likert scale. Mendelson and Aboud describe the

McGill Friendship Questionnaire-Friend's Function to uncover to what extent a friend fulfils certain expected qualities and the McGill Friendship Questionnaire-Respondent's Affection to test friendship satisfaction [Mendelson and Aboud, 1999]. Whilst designed for romantic relationships, the 7 point Relationship Assessment Scale proposed by Hendrick [Hendrick, 1988] was modified and utilised successfully by Bagwell et al. [Bagwell et al., 2005] to measure friendship satisfaction. Diary methods have been suggested as a less biased or intrusive way to gain self-report information.

Ickes claims that, at its best, observational data is as good as objective fact [Ickes, 2000]. He also extols the plethora of relationships that have been studied using observational methods to attest to its versatility as an evaluative method. However, the downside to this approach is the cost in time and equipment to record, rate, notate and analyse large amounts of data. Observational methods may also raise ethical issues with regards to privacy.

Tardy and Hosman suggest with experimental studies carried out in laboratories, you are presenting participants with unnatural or artificial situations and so your results may not generalise to the outside world. While this criticism may be valid for the study of already naturally occurring relationships such as romantic or father-son, when it comes to human-robot relationships there is little option but to craft the situation ourselves. The Robot House by LIREC demonstrates an attempt to combine laboratory-style control with the appearance of a real world, naturalistic environment.

Physiological data is another possible avenue for gathering information about close relationships. This type of information can be useful in its objectivity and in its visceral nature, but is not without its problems. Even though recent advances mentioned in Section 2.3.1 suggest the issues of expensive, obtrusive or invasive techniques are surmountable, it is still a daunting task to reduce this data and link it into an already complicated web of social theory [Ickes, 2000].

2.4 Conclusions and Research Questions

There are a multitude of fields that would contribute to the engineering of a robot capable of building a social relationship with a human based in musical activity. This chapter has surveyed the state-of-the-art these fields. It also covered the existing evaluative methods which one might use to determine the quality of any social relationship developed this way.

In this section, we summarise the challenges and gaps present in these research areas and propose three research questions that may address these challenges when answered. It will be these questions the research in this thesis addresses.

Whilst Section 2.1.4 recounted a number of HRI studies that go beyond a single session, they were mainly evaluated in the context of either functional tasks or therepeutic benefits. Indeed, throughout the literature in that section, Section 2.3.1 or Section 2.3.2, we failed to find an explicit model for a human-robot social relationship. Further, although one may expect that approaches to measuring human-human relationships could be transferred directly to measuring the quality of human-robot relationships, it is clear that there is too great a distinction between the former and the latter for this to be the case. This difference is attributed to the types and quality of social interactions afforded by current state-of-the-art robots. The reliance of this field and HRI in general on self report is also a concern. This leads us to the research question:

- **RQ1** How can we determine the quality of a social relationship developed between and human and robot over multiple social interactions?

This question will be addressed by the definition of a model for human-robot relationships. Favouring an approach of automated behavioural metrics, we will then explore which phenomena are indicative of the main facets of this model, namely, engagement, social presence and a close interpersonal relationship. HRI studies in this thesis will then demonstrate practical implementations of this methodological approach.

Section 2.1 covered research that pointed to humans being capable of engaging in social interactions, sometimes long-term, with non-humans. However, when the counterpart is a machine, this is often limited by the choice of interaction domain and design of morphology and behaviour in current state-of-the-art sociable robots. This often degrades further over time.

However, Section 2.2.1 shows ensemble music often facilitates social relationships. Further, Section 2.2.2 demonstrated that computers are capable of composing music in realtime in response to human musical input. Also, that robots are capable of playing acoustic instruments expressively. As such, it would be technically possible to build a robot that could improvise music with a human. Despite the clear potential to engage humans socially in this manner, existing academic studies exploring social connections through use of musical instruments are rare. Those that have been conducted are primitive, for example, music is reduced to non-simultaneous, single drum imitation game [Kose-Bagci et al., 2007]. This leads us the research question:

- **RQ2** Is it possible for music to provide the necessary engagement for a positive and sustainable social relationship between human and robot to develop?

We will first tackle this question by conducting a large online survey of human-human relationships. This uses well validated preexisting models to compare respondent's perceived provisions of the social relationship they have with a close friend and a regular co-musician. Comparisons of results for each category will demonstrate which facets are shared between the two relationships and show along which dimensions a positive, voluntary social relationship can be developed through musical interaction. Further, having used this information to develop a robot capable of improvising music with a human, long-term studies can be executed using the methodology developed in response to **RQ1** to examine if engagement can be provided and sustained.

Finally, any social relationship is made up of a number of social interactions and Section 2.1.3 showed that there is a number of behaviours necessary for any robot attempting a social interaction. It also demonstrated that adding social behaviours to a robot, such as an artificial personality and nonverbal behaviours, can increase engagement and social presence. This leads us the research question:

- **RQ3** Which social behaviours improve the potential for a positive and sustainable social relationship between human and robot based in music?

This question will be addressed by 3 HRI studies in which social behaviours are incrementally evaluated in controlled experiments where humans and robot improvise music together.

Chapter 3

A Model and Methodology For Human-Robot Relationships

In Section 2.4, we suggested that there was no directly transferrable methodology from either HRI or the study of human-human relationships that would allow us to determine the quality of a positive and voluntary human-robot social relationship. The need for this to be inplace before the subsequent research questions could be addressed lead to **RQ1** and it is with this question that this chapter is concerned.

We first define a model of the factors we deem necessary to be present in a human-robot relationship. A methodological approach to measuring these in the context of long-term HRI trials is then proposed.

3.1 A Model of Human-Robot Relationships

We have stated that a social relationship necessarily develops over multiple interactions, and that for these multiple interactions to be self motivated the experience must be an engaging one. However, the development and proliferation of complex technological devices is such that most humans in the developed world will regularly interact with machines that can provide them with near endless engagement and entertainment. These range from toys and computer games to word processors and social media platforms. However, whilst they may facilitate social relationships with other humans, most would not consider these as social relationships between human and machine in themselves. Below we detail factors that are the necessary additions to engagement for something to be classed as a human-robot relationship.

3.1.1 Social Presence and Believability

The concept of presence is often used to describe the illusion of non-mediation that can occur when having mediated experiences [Lombard and Ditton, 2006]. A subset of this is social presence or "the sense of being with another" [Biocca et al., 2003]. The prevalent occurrence of this is when humans communicate via technology, be it through a telephone, email, avatar or robot. However, social presence may also be used to examine the perceptions of socially intelligent artificial agents and in his definition of social presence as experiencing artificial social actors, Lee explicitly identifies social robotics as an example of this [Lee, 2004]. Being a psychological phenomenon, rather than a physical one, it is not concerned with the actual presence or not of another social being. Rather, it is concerned with the *sense* of being with one and as such it can be experienced when interacting with any animal, vegetable, mineral or indeed, robot. A social relationship must take place with another, and so social presence will be a necessary facet of any human-robot relationship.

Social presence has previously been used as a metric in HRI trials and research has shown a range of different variables may affect the perceived social presence of virtual or physically embodied agents. For example, in a study with synthetic voices in a book reviewing scenario, Lee and Nass demonstrate greater social presence when the personality of the voice is matched to the content of the text and the personality of the user [Lee and Nass, 2003]. They also find that an extrovert voice has more social presence than an introverted one. Lee et al. show that participant's feelings of social presence significantly affect whether they perceive a robotic pet with a complimentary personality more favourably [Lee et al., 2006b]. Further, Nowak and Biocca report a higher social presence from an agent represented by less anthropomorphic avatar in comparison to an agent represented by a more anthropomorphic avatar and one with no image [Nowak and Biocca, 2003]. They draw the conclusion that setting high expectations of an agent's faculties and failing to deliver can have adverse effects on social presence.

In related research, Aylett et al. suggest that there is enough evidence from Social Robotics research to imply the users believe systems are capable of having an inner state, goals and feelings [Aylett et al., 2011]. First mentioned by Sociable Robotics pioneer Cynthia Breazeal [Breazeal, 2004] and highly similar to social presence, they suggest a metric of believability and define it as the degree to which this user can suspend their disbelief and persuade themselves a robot possesses these characteristics. Believability is thought to increase engagement and can be aided by appropriate affective response.

Macdorman and Cowley continue on a similar theme by suggesting that a human body becomes a person by exploiting social mechanisms to create an identity for itself, so perhaps a mechanical body may also give the impression of personhood using the same devices [Macdorman and Cowley, 2006]? They report that a robot such as *Repliee Q1* that can replicate basic human attentional movements can give the impression of human presence.

The above demonstrates that social presence and closely analogous concepts are already being considered in the Social Robotics community. In combination with the fact it succinctly encapsulates the phenomena that elevate long-term engagement towards something that can be considered a positive and social human-robot relationship, it becomes an excellent candidate for inclusion in our model.

3.1.2 Behaviour in Social Relationships

Both verbal and nonverbal behaviour has been shown to be noticeably altered during multi-person interactions depending on the relationship that exists between people, their previous encounters and the current context. Although not universal, the meanings of these relational behaviours tend to hold well within particular social groups [Floyd and Erbert, 2003]. Within these, we are interested in behaviours which display a positive connection or liking between two people, or the the existence of a close relationship. In the context of our model, this means that if a positive social relationship exists between two parties, behavioural signifiers should be observable.

Of nonverbal behavioural cues, touch, proximity and posture are often the most revealing [Burgoon, 1991]. Burgoon demonstrated an open or relaxed posture was a sign of intimacy and informality in a relationship and that close proximity is a perceived sign of similarity and affection. She also found touch to be a sign of informality, trust and affection and that the type of touch was also important. For example, face touching portrays greater intimacy. Similarly, Noller reports that head nods, Duchenne smiles and forward leans are all common expression of love [Noller, 2006]. Another commonly seen set of behaviours are termed "immediacy" and tend to reflect engagement, liking and solidarity. Immediacy behaviours can include smiling, facial animation, body alignment and increased proximity. Guerrero and Floyd describe the signifiers of affectionate and warmth behaviour common in friendships as smiling, eye contact, head nodding and tilting and facial animation and expressivity [Guerrero and Floyd, 2006]. A forward lean is also seen as a sign of interest and facilitates interaction [Andersen and Andersen, 2004].

Although mutual gaze has been thought to be important in conveying immediacy, Abele reports that people tend to gaze away from each other during intimate moments in conversation [Abele, 1986]. This provides an interesting counter argument if we assume that those in a close relationship will experience a greater frequency of intimate moments. In concordance with this, Schulman finds an increase in "gaze-aways" over the course of a long-term study of counsellor-patient interactions. Schulman's studies also suggest that positivity is more important in early stages of a relationship, confirming earlier findings by Tickle-Degen and Rosenthal [Tickle-Degnen and Rosenthal, 1990]. The former reports fewer smiles and frowns in the later interactions and also draws links between decreased nodding over time and feelings of alliance [Schulman, 2013].

Mirroring or postural synchrony can also be a signifier of a close relationship and there has been much research linking posture to rapport [LaFrance, 1979]. Tickle-Degen separates this into *matching*, referring to similarity in body positions, *interactional synchrony*, referring to matching rhythms of behaviour and *mimicry*, referring to replicating a behaviour soon after the other has done so. All are presented as signifiers of rapport and positive interpersonal relationship [Tickle-Degnen, 2006]. Schulman found that posture shifts reduced during individual conversations and across multiple conversations, suggesting that these should be less prevalent during interactions between those in a close relationship [Schulman, 2013]. Whilst there is some evidence of posture shifts being related to discourse changes, rather than interpersonal feelings, his results were regardless of alignment with a topic shift.

Verbal behaviour can provide cues to relationships between people either through semantic content or delivery. Researchers have shown that people can accurately determine friends and strangers by articulation rate [Planalp and Benson, 1992], and that this tends to be faster in conversations friends or family members [Yuan et al., 2006]. Schulman demonstrates increased articulation rate, measured as reduced phoneme length, both as a conversation progressed and across multiple conversations [Schulman, 2013]. He suggests this demonstrates increased articulation rate as a signifier of interpersonal relationship.

Negative aspects of a relationship can also be identified through behavioural observation. For example, nonverbal behaviour can display dominance in a relationship and this is often influenced by the perceived power of the person [Dunbar and Abra, 2010]. Crick also identified relational aggression, as opposed to physical aggression, as an important factor when examining relationships. An example of this behaviour is deliberate exclusion from a friendship group as a form of retaliation or exerting control [Crick, 2009].

3.1.3 A Model of Human-Robot Relationships

Taking into account the above discussions, it is possible to succinctly define the factors necessarily present in a positive human-robot relationship.

A human-robot relationship can be defined as the development and maintenance of social presence and engagement over multiple interactions where the human party displays some behaviours indicative of a positive interpersonal relationship.

The scope, variance and individuality of social relationships is considerable and so any model which attempts to encompass all possible occurrences could be placed somewhere between ambitious and naive. However, the model above clearly lays out well defined, measurable concepts that may be used to begin to evaluate the existence and quality of a human-robot relationship.

3.2 A Methodology for Determining the Quality of Human-Robot Relationships

As social presence and engagement are key factors in the model described above, we will first cover existing approaches to their evaluation in a HRI context before presenting our preferred methodological approach to determining the quality of human-robot relationships.

3.2.1 Evaluating Social Presence

As the first to conceptualise it, Short, Williams and Christie's method for measuring social presence has been widely used [Short et al., 1976]. Following this example, the majority of researchers using social presence with artificial social actors use semantic difference scales or other self report Likert scales as a measure. The many approaches are detailed below.

Nowak and Biocca take 9 statements directly from Short, Williams and Christie's test when looking at interactions in virtual environments [Nowak and Biocca, 2003]. Another popular scale is the Temple Presence Inventory (TPI) constructed by Lombard and Ditton [Lombard et al., 2000]. This lists 103 items, some sections of which pertain specifically to social presence. Kasap and Magenant-Thalman take questions from the TPI to measure user responsiveness and engagement in the context of social presence [Kasap and Magnenat-Thalmann, 2011] whilst others develop their own scales from scratch. For example, Lee uses a 7 statement scale where users are asked about interactions with Sony's Aibo [Lee et al., 2006b, Lee et al., 2006a] and a 4 point scale with regards to a synthesised voice agent. Examples of questions are "While you were interacting with this Aibo, how much did you feel as if it were an intelligent being?" and "While you were interacting with this Aibo, how much did you feel as if it were communicating with you?". Whilst also using a questionnaire, Leite et al. also use video analysis to evaluate social presence [Leite et al., 2009]. They recorded events such as "user looking at iCat" and "user talking to iCat" as measures and then cross referenced these without questionnaire results. We believe this combination to be important as although we have established that social presence is concerned with a perceived experience and so can be measured by self report, it is questionable whether users have introspective access to this information. Also, as social presence may be something that fluctuates over time [Biocca et al., 2003], a post-hoc questionnaire may be too blunt a tool alone to capture this.

Another approach is to take other concepts from social psychology such as intimacy, immediacy or interactivity and use them as a marker for social presence [Biocca et al., 2003]. For example, Biocca et al. note that behavioural studies are common in face-to-face interaction research and suggest that a similar approach could be used for studying social presence, reasoning that if a particular social behaviour is observed, then the participant is experiencing social presence. An example of a similar experimental approach comes from Schermerhorn et al., who take well-established social-psychological concepts that are known to be affected by the presence of a human and suggest that the identification of these concepts in HRI trials imply that the human has perceived the robot as they would a human [Schermerhorn et al., 2008]. The effects they choose are that people are more truthful when interacting with computers and that performance of well rehearsed task is improved by the presence of another human, known as *social facilitation*.

As we saw with relationships in Section 2.3.2, making a one-to-one mapping between a physiological response and a socio-psychological one is inherently problematic.

3.2.2 Evaluating Engagement

Bickmore et al. define engagement as the degree of involvement the users chooses to have with the system and so use number and length of interactions as a sign of engagement [Bickmore et al., 2010]. They also claim that self reported intention to repeat interact can be a measure of a system's suitability for long-term engagement. Sidner et al. define it as 'the process by which two (or more) participants establish, maintain and end their perceived connection' [1], a definition which Castellano et al. deem suitable for describing social bonding between human and child [Castellano et al., 2012]. Both use a similar post hoc self report measure of engagement.

In a separate paper, Castellano et al. suggest engagement with a robot may entail an affective and attentional component based upon positive feelings and a willingness to maintain interaction [Castellano et al., 2009]. They report a system capable of classifying engagement with 94.79% accuracy when using

[1] [Sidner et al., 2005]

both contextual information and nonverbal cues. Indeed, gaze, speech and body language have all been identified as important in displaying engagement in HRI [Ivaldi et al., 2015]. Nonverbal cues used in the above study included the user smiling and the user looking at the robot. Additionally, Sanghvi et al. describe a system capable of automatically detecting engagement from laterally captured posture data with an accuracy of 82.2% [Sanghvi et al., 2011].

3.2.3 Methodological Approach

Of the 3 research questions proposed in Section 2.4, the first deals directly with how the quality of a human-robot relationship may be evaluated. Developing such as a method is a prerequisite for answering the following 2 research questions. Having encapsulated a human-robot relationship into a model in Section 3.1.3, we now outline a methodological approach for measuring the phenomena it covers. In doing so we go some way to answering **RQ1**.

Critical of the post-hoc self report approach taken by many HRI researchers, we propose the automated, multimodal measurement of behaviour in controlled experiments as the favoured approach for the evaluation of engagement and social presence in long-term HRI trials. Outside of talk-through approaches used in HCI, self report will be necessarily after the fact, providing a low resolution measure of an interaction that is removed from its original context. Further, even if actions conducted and the reasons for the conducting og these actions were accessible to a participant introspectively, and they could express this on a Likert scale or equivalent, the social bias against having a relationship with a robot is likely to skew results. Survey-based self report is also an issue in long-term trials as there may be a learning or boredom effect through repeated measures.

As an alternative, we detail several social behaviours that commonly occur in relationships and posit that single occurrences or changes over time can be taken as signifiers of a relationship between the two. Further, we propose that insights into the engagement of a participant and the social presence they experience can also be gained from behavioural observation. For example, time spent together voluntarily is a display of the former and signifiers of treating the robot as they would a human for the latter. By recording these a high resolution, largely repeatable, temporally accurate measure is gained that is unbiased by what the participant wishes to disclose.

We propose these measures should be multimodal, as face and the body movements have been shown to be critical for communicating behavioural state. Further, the automation of the data collection process is key, as it allows us to generate the large datasets of these high resolution multimodal behavioural measures during long-term studies in a relatively cheap and efficient manner. Automation also ensures continuity across sessions, ensuring scientific rigour when repeatedly measuring data throughout long-term studies.

One of the most important things to keep in mind is that the presence and strength of any phenomena chosen will only be indicative of engagement or social presence if it can be reliably deemed so in the context of the interaction. For instance, close proximity has been identified as a indicative of interpersonal relationships, yet if any experiment involves an immobile robot and a seated participant, this measure is not a useful one. Similarly, although some have suggested posture synchrony as a sign of rapport, if the robot has significantly less degrees of postural freedom than a human, this phenomenon is unlikely to be observed. In short, due to the lack of universality within the functionalities, use and context of social robots, it is unlikely that any specific metrics will hold true across all scenarios, indeed, it is quite likely some metrics will not be transferrable because phenomenon simply does not exist from one context to the next. For example, a measure of the amount of tips a bartender robot receives in a shift could not be used in a teacher pupil situation. It is up to the researcher to pick appropriate phenomena they deem to demonstrate either engagement or social presence or that provides evidence of an interpersonal relationship between the two.

Moreover, it is important to consider how suitable a phenomenon is for automated measurement, taking into account cost, obtrusiveness, reliability and accuracy. For example, button presses on a keypad can be logged with millisecond accuracy and almost complete reliability. The storage of this data as text files or databases is also inconsequential. However, measuring facial expressions from video footage requires it to be synched to any other data streams, researching and implementing a preexisting algorithm or the developing of your own if necessary followed by computationally expensive processing to get data. Even then it may suffer from misclassification errors and incompleteness depending on your choice of algorithm and the quality of your corpus.

Chapters 6, 7 and 8 will describe a practical implementation of this methodological approach in HRI experiments involving our robotic drummer *Mortimer* and a human pianist. As all 3 studies have broadly the same set up, many of the measures will be applicable to all conditions in all studies, however, the idiosyncrasies of some situations may require slight variations in the measures and the statistics used to analyse them.

Chapter 4

The Provisions of Human-Human Musical Relationships

4.1 Introduction

In Section 2.2.1 research was presented that extols the potential for music to provide a great opportunity for shared creativity and social activity. This, along with the failure of other robots to maintain engagement over multiple interactions lead us to **RQ2**. This asked whether it was possible for music to provide the necessary engagement for a positive and sustainable social relationship between human and robot to develop. Further, in **RQ3** we raised the question of which social behaviours improve the potential for a positive and sustainable social relationship between human and robot based in music. It is these two research questions that this chapter addresses.

As such, to further investigate the possibility and nature of human-robot relationships centred around musical activity we used the NRI [Furman and Buhrmester, 1985] survey to compare respondent's friendships, used as an example of a positive, voluntary social relationship, with their relationships with a regular co-musician. We administered the original version of the survey, known as the NRI-SPV, as it is heavily influenced by Robert Weiss and Harry Stack Sullivan's widely used models of what a social relationship may provide [Furman and Buhrmester, 1985]. Similarities were found between a key set of perceived provisions of relationships based in regular musical activity and friendships, namely reassurance of worth (WOR) and instrumental aid (AID). This provides evidence that a relationship based in regular musical activity can provide similar facets to a friendship and so can itself be a solid grounding for building positive, voluntary social relationships. In doing so, it points towards an affirmative answer to **RQ2**. Further, these findings allow us to outline design specifications for developing a robot and composition algorithm to best build relationships through joint musical activity and point to social behaviours which may help when answering **RQ3**.

4.2 Friendship

Friendship is an often close bond which takes time and high-level skills to initiate, maintain and develop. Although there have been many examples of family bands, of duetting couples, and of band members who were definitely not friends, friendship was chosen as a comparative relationship in this survey. This is not because we believed that we would be able to recreate such a tight bond between human and robot within this research, but because it is a relationship descriptor that will be readily understood by the majority of, if not all, participants. It is also a relationship a large proportion of respondents will be engaged in currently. More importantly, we chose it as it was, amongst the widely recognised, widely experienced and culturally consistent relationships, the one we predicted would be most similar to the relationship that may eventually develop between human and robot through regular musical activity. Below we present the case for friendship as a better fit for our purposes than romantic and kin relationships.

In comparing friendships and romantic relationships, Wright notes that friendships tend to be less exclusive [Wright, 1984]. This means that we will not have to deal with jealous spouses when conducting our HRI studies. By proposing human-robot relationships we are already challenging some quote strongly held social norms and so it is important to pick a relationship which will minimise the effects of this. Friendships aid us in this respect as they are less regulated by social rules. Romantic relationships also tend to be more brittle, whereas friendships tend to survive setbacks better [Derlega and Winstead, 1986], helping us avoid early conflict. On a practical note, Laursen notes that physical intimacy is the main difference between friendships and early romantic relationships [Laursen and Pursell, 2009], however, the engineering involved to fulfil the physical commitments of a romantic relationship may go beyond the scope of this research.

With regards to kin, although all apart from the least fortunate of us will have experienced at least one living family member, there can be wide variance in the relationships between family roles. For example, the relationship that would exist between a father and son or between siblings can differ significantly. There will also be wide variance between one father and son and another father and son. Even if a specific relationship within a family was picked, if it is intergenerational it may be that cognitive differences between age groups result in both parties are actually not in the same relationship [Acitelli et al., 2000]. This complication is avoided by the norm of age similarity of friendships. Friendship is also a voluntary bond [Wiseman, 1986], unlike a kin relationship, so in this respect is a closer match for our purposes.

4.3 Method

Originally developed to be able to examine perceived characteristics of a person's relationships with those in their social network, the NRI has been requested for use by over 900 researchers since 1985. There have also been two publications [Furman, 1996, Furman and Buhrmester, 2009] that attest to its validity as method for measuring participant's perceptions of their relationships.

Our survey asked participants to pick two distinct people, one who was considered "a good friend" (F), and one who they played music with regularly (M). They were then asked to fill out the the NRI-SPV about each. The NRI-SPV provides scores for 9 subscales, each pertaining to a relationship provision. Derived from the work of Robert Weiss, these are affection (AFF), instrumental aid (AID), reliable alliance (ALL), antagonism (ANT), companionship (COM), conflict (CON), intimate disclosure (DIS), nurturance (NUR) and reassurance of worth (WOR). Each subscale is calculated from 3 distinct questions, of which the participants must rate how much they feel a statement relates to the relationship they are describing on a 5 point Likert scale. These provide individual scores that can also be aggregated further into complete positive and negative scales. Of the 9 subscales, 7 represent positive qualities and 2 represent negative qualities. Example questions are displayed in Table 4.1.

Table 4.1: Provisions Addressed by NRI-SPV with Example Questions

Provision	Example Question
AFF*	How much does this person really care about you?
AID*	How much does this person teach you how to do things that you dont know?
ALL*	How sure are you that this relationship will last no matter what?
ANT**	How much do you and this person get on each others nerves?
COM*	How often do you and this person go places and do things together?
CON**	How often do you and this person argue with each other?
DIS*	How often do you tell this person things that you dont want others to know?
NUR*	How much do you take care of this person?
WOR*	How much does this person treat you like youre admired and respected?
* positive,	** negative

Respondents were invited to take part via email and internet message boards. Recruitment was aimed primarily at university music societies in the UK and US. Adverts were also posted on message boards for both professional and amateur musicians and on websites providing services for introducing musicians to each other and employers. Participation was voluntary, but a prize draw of vouchers was offered for those who did take part. The final number of respondents (n=141), represents a reasonable return. The gender split was approximately equal, (Female n=69, Male n=73). For the age of participants, an overall range of 18-67 and a median of 20 represents a strong skew towards young adults. This was to be expected on account of the main distribution targets.

4.4 Results

In order to determine whether the participants perceived their relationship with F and M differently, the scores for F and M for the positive and negative scale and each individual subscale were subjected to non parametric Wilcoxon signed rank tests. Results are displayed in Table 4.2. After testing the internal reliability of the questions that built up each subscale, we report a highly satisfactory mean Cronbach's Alpha of .90. A Cronbach's Alpha across all positive subscales of .90 for M and .89 for F demonstrates the positive scale as a good generalisation of subscale scores.

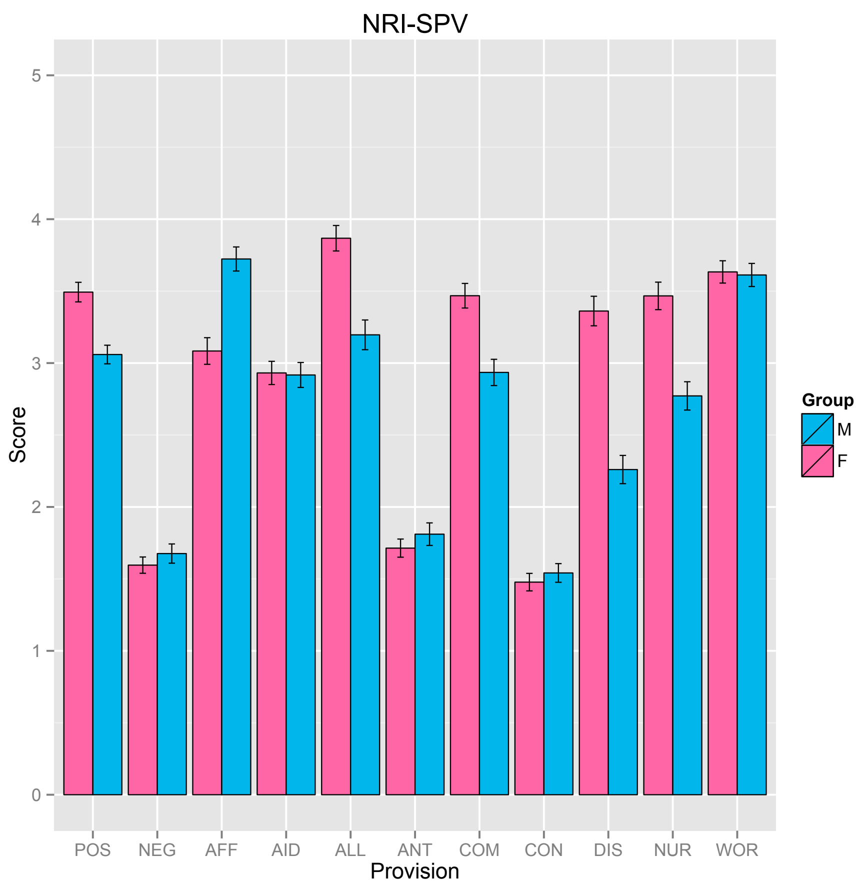

Figure 4.1: Mean Scores for M and F for Online NRI-SPV

Table 4.2: Wilcoxon Signed Rank Test Results for M and F for NRI-SPV

Pairs	*Mdn* (F)	*Mdn* (M)	*z*	*r*
POS (F) - POS (M)	3.57	3	-5.59*	-0.47
NEG (F) - POS (M)	1.33	1.33	0.78	0.07
AFF (F) - AFF (M)	3	4	5.61*	0.47
AID (F) - AID (M)	2.67	3	-0.28	-0.02
ALL (F) - ALL (M)	4	3	-5.23*	-0.44
ANT (F) - ANT (M)	1.33	1.67	1	0.08
COM (F) - COM (M)	3.67	3	-4.3*	-0.36
CON (F) - CON (M)	1	1.33	0.68	0.06
DIS (F) - DIS (M)	3.33	2	-7.25*	-0.61
NUR (F) - NUR (M)	3.67	2.67	-5.39*	-0.45
WOR (F) - WOR (M)	3.67	3.67	-0.64	-0.05

* $p<.0001$

4.4.1 Difference

Weiss reports that relationships tend to specialise and so any single relationship is unlikely to impart all provisions. As such, we do not predict all provisions will be perceived as being imparted similarly by F and M. In some cases, this will be for practical reasons based on the nature of the interactions and this was seen with intimate disclosure (DIS) and nurturance (NUR). Results demonstrated that these not only showed the most significant differences between F and M, but also reported the lowest mean scores for M (See Figure 4.1). The opportunity for self disclosure is noted as a key provision of a social relationship by Weiss, as well as in many subsequent studies, for example, that conducted by Lowenthal and Haven [Lowenthal and Haven, 1968]. However, there is no clear way to provide this through a relationship centred in musical activity as it relies mainly on verbal communication. Nurturance (NUR), the opportunity to look after another, also scored significantly higher for F than M. Again, this is a provision that the interactions partaken with M does not intuitively afford and one that is usually related with family relationships, especially infant-caregiver [Weiss, 1974, Furman and Buhrmester, 1985].

Reliable alliance (ALL) is also closely linked with kin relationships [Weiss, 1974] and is characterised as a lasting and dependable bond. F was perceived as a significantly higher provider than M for this subscale and this seems consistent with the research that views friendships as robust and long-term, whereas there may not be the commitment to maintaining a musical relationship beyond a certain project. This notwithstanding, although being significantly lower than

F, reliable alliance (ALL) had the second highest mean score for M (Figure 4.1). Duck reports that friends can provide a sense of inclusion through group membership and clearly this is something provided by M as well, if to a somewhat lesser extent [Duck, 1991] .

Companionship (COM) is linked to social integration. Clift and Hancox cite that 85% of singers in a community choir reported social benefits from singing together [Clift and Hancox, 2001]. Likewise, just as friendships can give a sense of social inclusion by providing someone to share private understandings or languages with [Duck, 1991], regular musical partners learn each others styles and preferences over time, especially in an improvisatory setting. This seamlessness allows for more of the highly satisfying experiences that Csikszentmihalyi reports to be desirable in creative activities [Csikszentmihalyi, 1997]. However, scores for F were again significantly higher for companionship (COM), as was also the case for affection (AFF). This is related to emotional attachment and whilst you may grow to like someone you are playing music with, it is unsurprising that a good friend is perceived as a greater provider of this factor.

4.4.2 Equivalence

Both reassurance of worth (WOR) and instrumental aid (AID) showed no significant difference between the scores for F and M. However, this does not allow us to claim equivalence between F and M on these provisions. Using the procedure developed by Tryon and Lewis [Tryon and Lewis, 2008], the inferential confidence intervals were calculated for both groups and then claim statistical equivalence if the difference between the minimum lower bound and maximum higher bound is within a minimum difference deemed inconsequential enough (δ). For this a δ of 0.25 (5%) was chosen. It is worth noting that the decision to apply this test, and so the δ value, was taken after the experiment when seeking to further investigate the close proximity of scores for WOR and AID for F and M. However, we believe 5% to be a fair boundary and regard this analysis as a valid conclusion of equivalence. The results for this are demonstrated in Figure 4.2 and Figure 4.3.

Reassurance of worth (WOR), defined as how much a relationship affirms ones feelings of competence or value, showed statistical equivalence (δ=0.25) and had the highest mean score for any of the subscales for M (See Figure 4.1). This similarity aligns with research that reports friendship as a reinforcer of personality and beliefs [Duck, 1991] and that which links shared musical

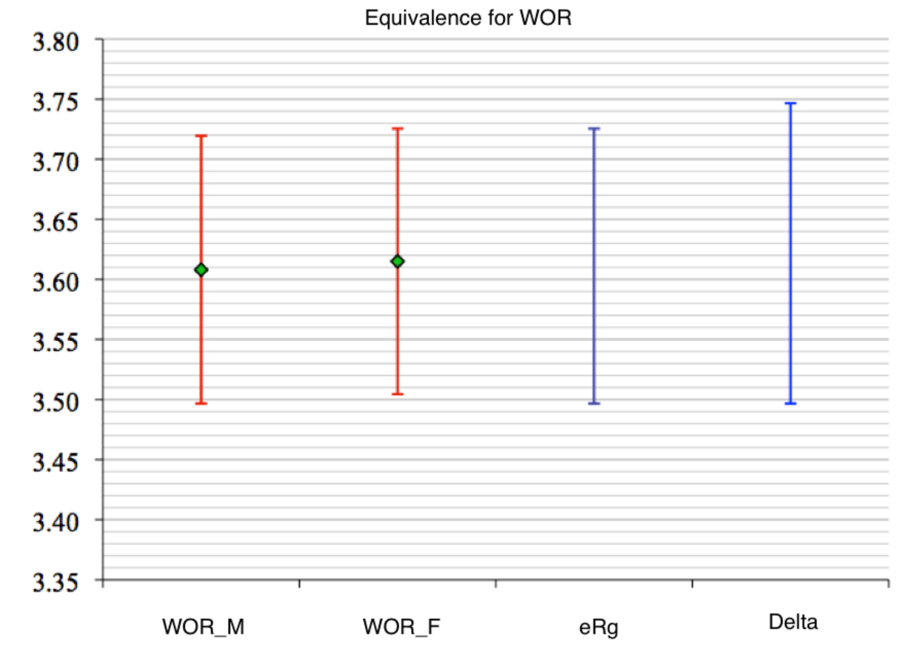

Figure 4.2: Equivalence between M and F for Reassurance of Worth (WOR)

preference to shared values [Boer et al., 2011]. It also relates to the results seen by Kokotsaki, whose music student respondents relayed strong feelings of usefulness to the group and a shared pride in success whilst playing in ensembles, particularly those of a small size [Kokotsaki and Hallam, 2007], and to those seen by Clift and Hancox, 75% of whose participants also reported emotional benefits from singing in a community choir [Clift and Hancox, 2001]. Further, Duck believes that because friendship is a voluntary relationship, as playing music together often is, the mere fact that the other has chosen to spend time with you is a boost to self-esteem [Duck, 1991].

Instrumental aid (AID) or guidance is the other positive subscale to show equivalence between F and M and a great amount of progression, personal development and co-learning can result from musical activity. Tangible provisions such as this are regarded as an important part of human sociality and one of the main advantages of socialising for humans is the exchange of knowledge and skills. It has also been proposed that a friendship can be charaterised by the provision of services that cannot be readily purchased by other means, in the way that, for example, physical goods can [Duck, 1991]. It is this type of experiential guidance that is often imparted while playing music together.

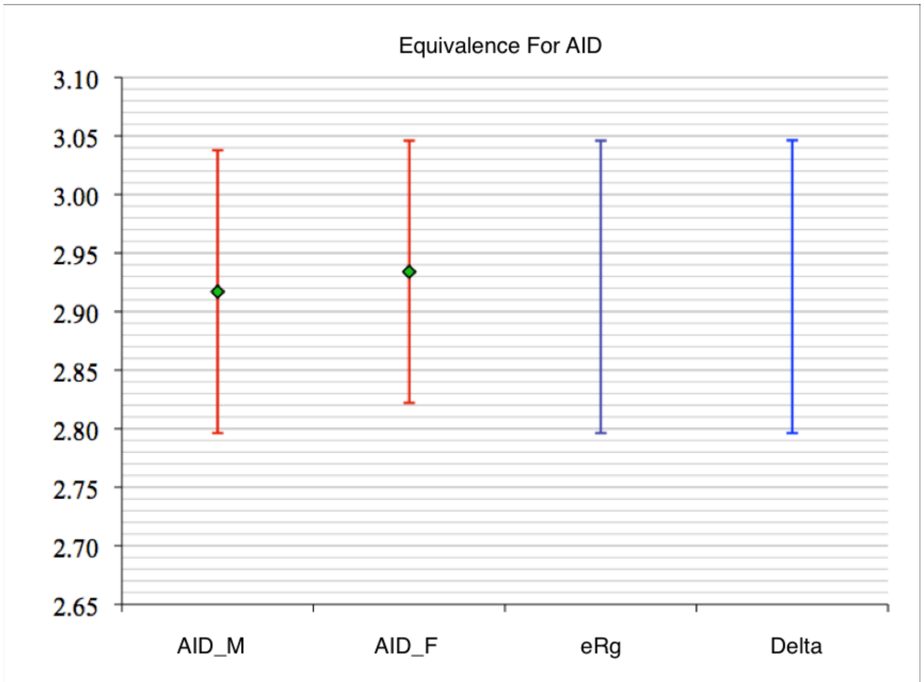

Figure 4.3: Equivalence between M and F for Instrumental Aid (AID)

4.4.3 Effect of Standard and Regularity

Using linear regression, we found that the standard to which the respondents played with M, rated on a 5 point Likert scale from beginner to expert, significantly predicted the positive score for M, b=.17, t(140)=2.04, $p<0.05$. It also explained a significant amount of the variance in positive scores $R^2=.02$, F(1,140)=4.16, $p<0.05$. When tested against the individual positive subscales, the standard played was only a significant predictor for reliable alliance (ALL), nurturance (NUR) and reassurance of worth (WOR). These findings suggest a frustration at a partner's or one's own ability when at playing to a lower standard. They can also be explained by Weiss, who reports that reassurance of worth is commonly seen amongst colleagues in highly skilled or difficult roles [Weiss, 1974]. Indeed, the ability to succeed in any task which is perceived as difficult by others is one which will raise self esteem, so if a relationship is integral to this achievement, it is likely to be rated highly. Csikszentmihalyi describes the balance of skill and challenge as integral to maintaining optimal experiences, especially in creative situations [Csikszentmihalyi, 1997]. Conversely, how often people play together could not significantly predict either positive or negative scores.

4.4.4 Effect of Age and Gender

The majority of participants chose a person for M within 3 years of their own age (N=108), with an even higher portion choosing a similarly aged person for F (n=122). However, absolute relative age was not a significant predictor of positive or negative score for M, yet it was with positive scores for F, b=-.345, t(140)=-4.33, $p<0.001$. Relative gender was also not a significant predictor of positive or negative scores for M.

4.5 Design Implications

Beyond supporting **RQ2** by demonstrating further that friendships and relationships involving regular musical activity share some key aspects, the purpose of this study was to inform us in what particular ways these relationships are similar. If we are to answer **RQ2** and **RQ3**, then insight into the social and musical behaviours to be included in a robot to best promote a positive social relationship with a human is invaluable. Using the detailed information about perceived provisions from the NRI-SPV and knowledge of what people expect to gain from a musical relationship, we now consider the implications for the design robot to achieve such ends.

In Section 4.4 of this chapter, we discussed reasons why we thought that participants perceived musicians as good reassurers of worth. This is now related back to our overall research as considerations are made how to design a robot to reassure worth well. Primarily, this can be achieved by explicitly reassuring a human of their skill through vocalisations. For example, flattering them when they succeed and not admonishing them when they fail. Although there may be fears that artificially generated praise will seem disingenuous, Reeves and Nass have shown flattery was appreciated when coming from a computer, even if it was unwarranted [Reeves and Nass, 1996].

A less explicit approach in reassuring the user of their competence will be the balancing of difficulty and skill. Also identified by Csikszentmihalyi as an important condition for a *Flow* experience, there are two possible outcomes that can occur if the difficulty of a task and the participants skill level are not equally matched [Csikszentmihalyi, 1997]. If the challenge is too great then the participant will become anxious or frustrated. Conversely, if the challenge is not enough then the participant will become bored with the simplicity of the task. Either will undermine the users feelings of competence or worth. This balance is also not a static state that once found may be left. As the participant repeats the activity, their skill level will increase, and thus the complexity of the challenge must increase proportionally.

With respect to an IMS, designing a system that provides this balance, even for a known performer, is challenging. The presupposition that you are able to decide a performer's skill level and then match this to a "challenge" level in your system involves many assumptions. It would appear that any system, especially one that will be used by more than one user, would have to be able to detect whether the user was struggling, or indeed was becoming overcome with ennui, automatically and then adjust itself accordingly. The former would be easier to detect, for instance, irregular timings, perhaps often behind the beat, may be the result of the performer becoming lost and confused. However, there is little way of telling whether this is actually the case or being creatively for expressive effect. Long silences may be another identifier, although waiting until your co-performer has become so frustrated they have given up completely does not bode well for blossoming friendships. Ideally, detection and repair would occur earlier. Allowing the user to explicitly signify when they have become lost may be a solution to this.

Noticing when the user has become bored is more of a challenge, yet may pose less of detrimental effect during the interaction. Often, in a dyadic musical improvisation, one participant will keep a steady pattern, allowing the other to be more free and experimental. These roles are often traded throughout, allowing each participant to take part in either. In Chapter 5 we will describe the development of a robotic drummer. As this robot does not have to listen to the human to keep time, in that it will be the one keeping the beat, as is the norm for a percussion player, the human does not have to worry about moderating the complexity of their play in order to not confuse the robot. Thus in theory, they can always be playing at the edge of their ability and so should not become bored in this respect. However, this is not to say that the optimum experience will occur through providing a simple track for the human to jam over.

The robot should provide enough variance and new ideas to inspire novel responses by the human, and if this is not forthcoming then the human may indeed become bored, feeling as if they are taking all the cognitive load of the improvisation, or even loosing the sense of a collaboration entirely. There is also nothing to say that this balance must come through a seamless automatic procedure. Two humans may well vocalise their feelings towards the musical outcomes during pauses in a session. As such, explicit control between interactions from the human to robot on how they would like forthcoming interactions to proceed could be easily implemented and go a long way towards balancing skill and challenge in a manner that could vary over time. In summary, the sensing of difficulty or boredom need not necessarily be explicitly enacted, but the damaging effects can be allayed by creating a situation where the user is always able to play uninterrupted and at the edge of their ability whilst being provided with novel and related content to react to.

The other provision to show equivalence between M and F was instrumental aid (AID). The obvious way to work this into our system would be the adoption of tutor-student roles, however, the provision of guidance could be implemented in a less formal or explicit way. By the very nature of providing a platform for the user to play along with a percussion accompaniment, the user will develop their skills for duet performance. Pianists often practise with a metronome to hone their skills at playing in time, however, an IMS could provide a much more engaging interaction. Beyond helping them keep in time, playing with a robotic drummer will improve their ability to improvise alongside percussive accompaniment. Further, if the robot is reactive to their playing, they will learn how to react to each other's cues in realtime.

Other advantages that come with using a robot is that it never gets tired, is always available and most importantly does not judge. Especially for beginners and intermediates or anyone trying to learn a new skill, the feeling of self-consciousness at lack of ability can be a barrier to seeking out others to play with to develop skills of playing in ensembles. This should not be the case with a robot. The robot can provide a platform to build up confidence for ensemble performance by giving a situation absent of judgement and so aid feelings of competence and provide the opportunity to improve.

4.6 Conclusions

Using an online study, we found that friendships were perceived as significantly better sources of a number of positive provisions of a social relationship than a relationship with someone you play music with on a regular basis. However, there were near equivalences for a key group of provisions, suggesting that these provisions of a friendship may also be rendered by a regular co-musician. Namely, these were reassurance of worth (WOR) and instrumental aid (AID) or guidance.

Whilst being wary of drawing direct parallels from studies involving human-human relationships to human-robot relationships, we conclude that, as expected, regular musical activity alone is not a necessary or sufficient condition for building complete human-robot relationships. However, it may impart some key provisions for moving towards this goal in a way others have failed to do so. In relation to **RQ2**, this suggests that it is possible for music to provide the necessary engagement for a positive and sustainable social relationship between human and robot to develop. Further, this knowledge gives vital cues for designing sociable technologies to engender human-robot relationships based in music when further investigating **RQ3**.

Chapter 5

Technical Development

The research in Section 2.2.1 and Chapter 4 has provided strong evidence in support of **RQ2**. The next step in investigating the appropriateness of musical activity as a provider of engagement is further study with live HRI trials. This will require us to build a robot capable of playing music with a human participant. The necessity to develop a robot to take part in HRI experiments is also crucial to addressing **RQ3** and as such any platform should allow for the incremental addition of simulated social behaviours. Although WoZ approaches can be useful in the early prototype stage of HRI trials, the robots that people outside of the academic environment will interact with in the near and distant future will have to run autonomously to some extent and so experiments with operational technology should be the priority of the field.

The development of musical robots as physically embodied accompanists, performers, improvisers or art installations is not unique. However, we stated the primary focus of the research in **RQ2** and **RQ3** as investigating the development of sustainable and meaningful relationships between human and robot based on regular, open-ended musical interaction. As such, we have endeavoured to build a robot to best match these aims.

Design implications drawn from our results in Chapter 4 were also taken into account. These informed us that a good balance of skill and challenge and providing the opportunity to learn are both key parts of a friendship based around musical performance. Following this, we identified three key characteristics that have been balanced during development to maintain the long-term engagement and elicit the feelings of social presence we have stated as necessary facets of a human-robot relationship.

The system must be stable enough to be run unassisted by a novice user, as any malfunction or erratic behaviour will diverge from the positive experience needed for engagement and break the illusion of believability that creates feelings of social presence. There are a number of competencies, research fields in their own right, that are necessary for even a basic musical interaction. Arguably, if these are not met to a sufficient level, any findings will have to be moderated, in that, a negative response that could be interpreted as a function of the interaction on a social level could not be separated from frustration at lack of musical ability. To avoid this confusion, tools from ongoing but unsolved research areas such as online beat tracking and Automatic Speech Recognition (ASR) have been avoided.

It is also critical that the robot is responsive to human playing. If the relation between the input and output is explicit enough to be recognised by the player then this will give the impression the robot is listening to them. If the response seems intelligent and adds to the playing experience positively, this will at worst demonstrate the system's competent engineering and add provide an enjoyable experience to the listener. This is likely to increase long-term engagement. At best, this will addi to the feeling of social presence and layi the foundations for a more meaningful relationship by giving the impression that the robot has some understanding of the music being played and its emotional content.

Dannenberg et al. suggest that to reach the most users the future of human computer musical performance systems should have a "focus on steady-tempo popular music" [1], as opposed to inhabiting the experimental music field much research occurs in. Acknowledging that the pool of potential beneficiaries of this research has already been narrowed by picking the domain of musical performance, we are aiming to provide a system which can be accessible to as many as possible. As such, when selecting composition styles and instrumentation, we should pick a musical context that accommodates a broad audience.

Developing novel robots for each HRI trial is a non-optimal solution as it leads us as a community away from developing a solid body of comparable

[1] [Dannenberg et al., 2013]

studies. However, due to the accuracy and synchronisation required in robotic musicianship, a bespoke system was considered necessary.

We now outline the construction and software development of *Mortimer*, a robot capable of playing an acoustic drumkit and responding in realtime to human piano input expressively and musically. The physical robot, all electronics, the control systems and composition algorithms were developed by the author from scratch unless stated otherwise.

5.1 Instrumentation

Mortimer, shown in Figure 5.1, is built around an aluminium frame and with two identical beater arms, with the intention to trigger slight anthropomorphism whilst still presenting a robot which is clearly mechanoid. The beater arms are made from clear acrylic and are powered by 12V 7W pull solenoids. To strike the instrument, the back end of the pivoted beater is pulled upwards, pushing the striking end down. At the end of the strike, the solenoid is turned off and the beater is reset by gravity. If the the strike time is less than the time taken for the solenoid to reach its maximal pull length, the strike time determines how far the beater moves down towards the instrument. Once the solenoid has reached its peak, a longer strike length determines the time the beater stays against the surface of the instrument. Both of these effect the velocity and timbre of any individual strike, as well as the retrigger rate possible for each beater. A shorter strike time allows a faster retrigger. This parameter provides a great deal of variation and expressivity and so must be calibrated exactly for each instrument. If two instruments are triggered from the computer at the same time but take differing times to make contact and so sound, then differing strike times also have bearing on the synchronisation between instruments. To counter this, each arm also has an offset to ensure synchronicity between instruments. Figure 5.2 shows a profile of the arms and beater mechanism

As well as the two striking arms there is an automated kick drum, the mechanism for which is displayed in Figure 5.2. Made from a modified kick drum pedal, the beater presents a different engineering challenge to the horizontally mounted snare and hi-hat. The kick was placed on its side, to conform with a standard drum kick set up and a simple pivot could be used to ensure the return of the solenoid through gravity alone. A rack and pinion was used allow

Figure 5.1: Photographs of the Development of the Drumming Robot, Mortimer

the most effective transition from the linear movement of the solenoid to the rotary motion of the axel.

5.2 Composition

An algorithmic approach to generating the content for *Mortimer* has been chosen. This gives the advantages of avoiding the possibly repetitive nature of playing back a corpus of already composed music and as long-term engagement is sought, this variation is paramount to avoid mundanity. Further, it does not require a large corpus of drum rhythms to be composed by hand, which in itself would be a time consuming process. It also allows the participant's own playing to be taken as an input to the composition process, thus providing real-time response to their playing. We predict that a drummer that can react appropriately to a user's musical input has a greater chance of providing long-term engagement as the user explores the interplay between the two, rather just their own accompanied playing.

The proceeding section describes the algorithm used to compose the scores that *Mortimer* plays, the measures taken from the user and how these influence the former. Figure 5.3 demonstrates the system in its entirety.

5.2.1 User Input

Explicit User Input

In the discussion of the design implications provided by our research in Chapter 4, we cited that the balance of challenge and skill was crucial for reassuring the the user of their competence. We reported possible automatic processes to achieve this but also the challenges faced with implementing them and as such have allowed the user to explicitly change some performance parameters to influence the composition. Between human musicians this may happen verbally, although as there is no ASR implemented in the system a tablet interface with sliders has been used to facilitate this. The parameters specified are tempo, complexity and length.

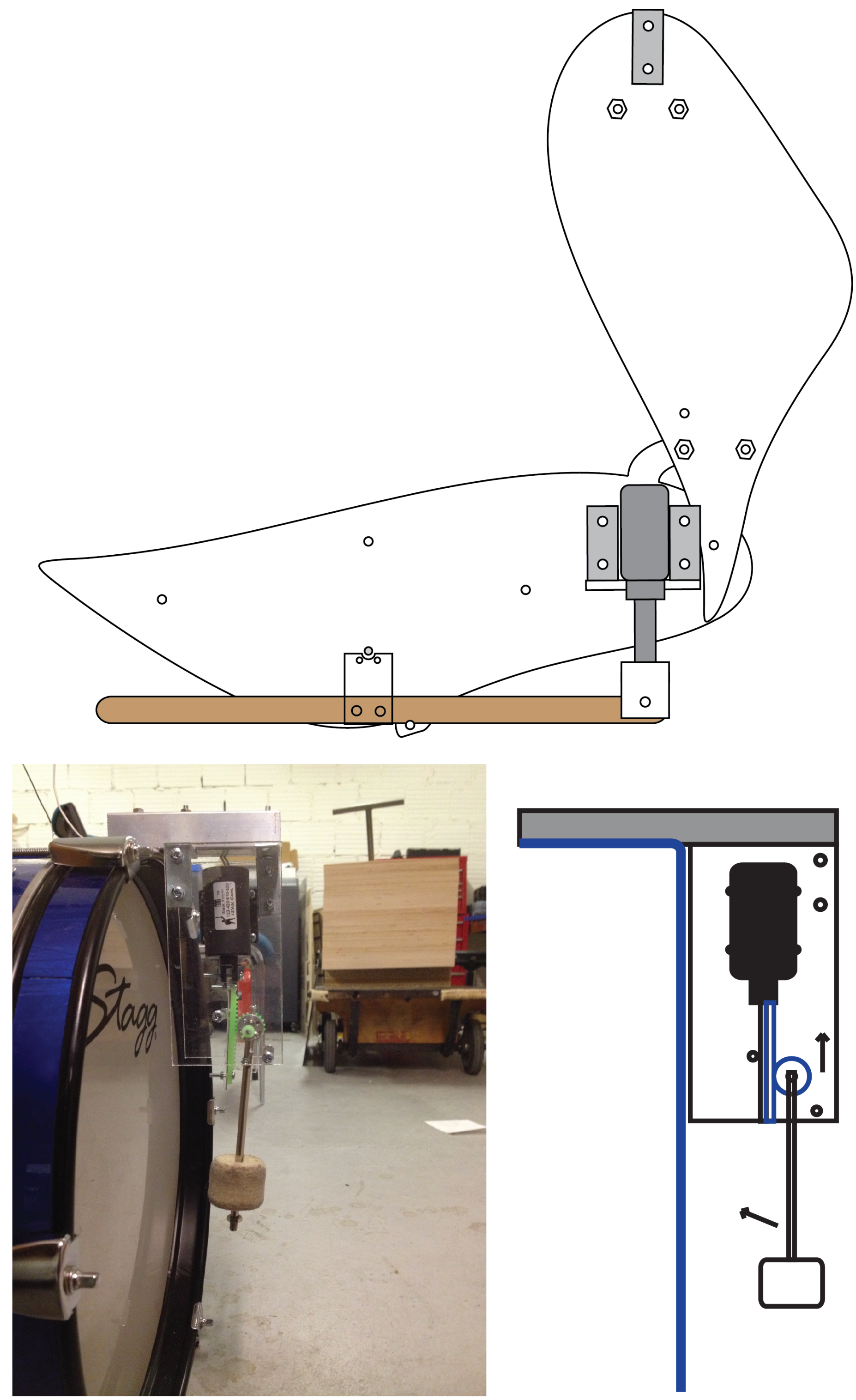

Figure 5.2: Technical Drawings of the Arm and Kick Mechanisms for the Automated Playing of Acoustic Drums

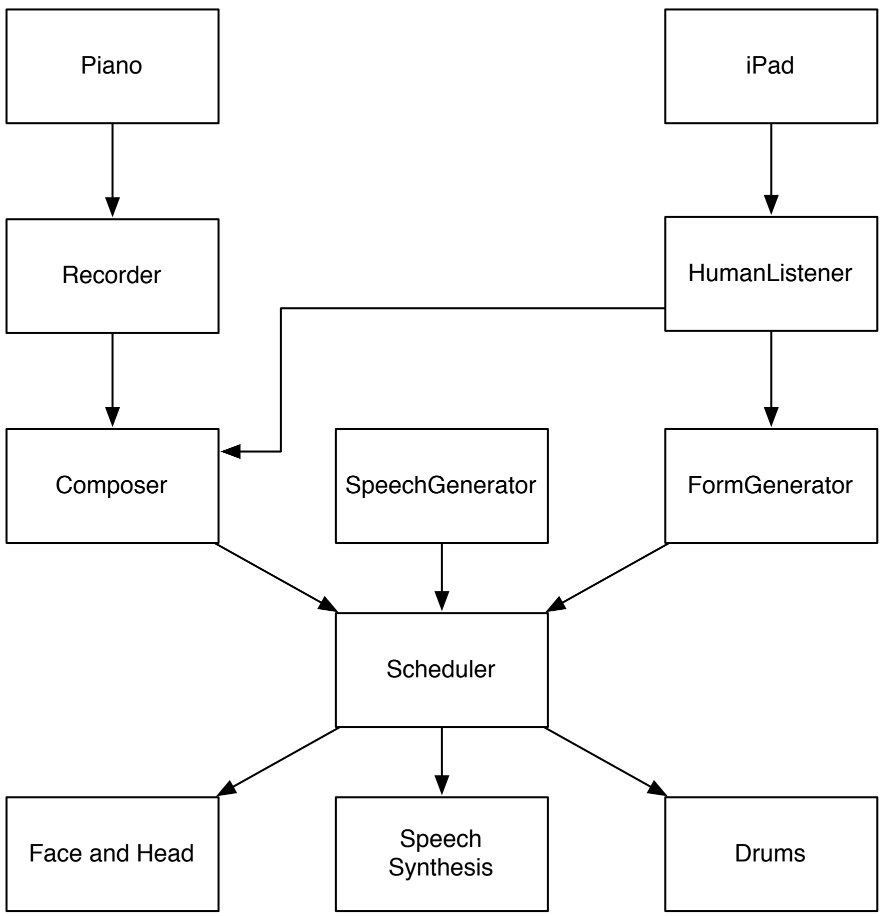

Figure 5.3: A Diagrammatic Overview of Mortimer's Composition System

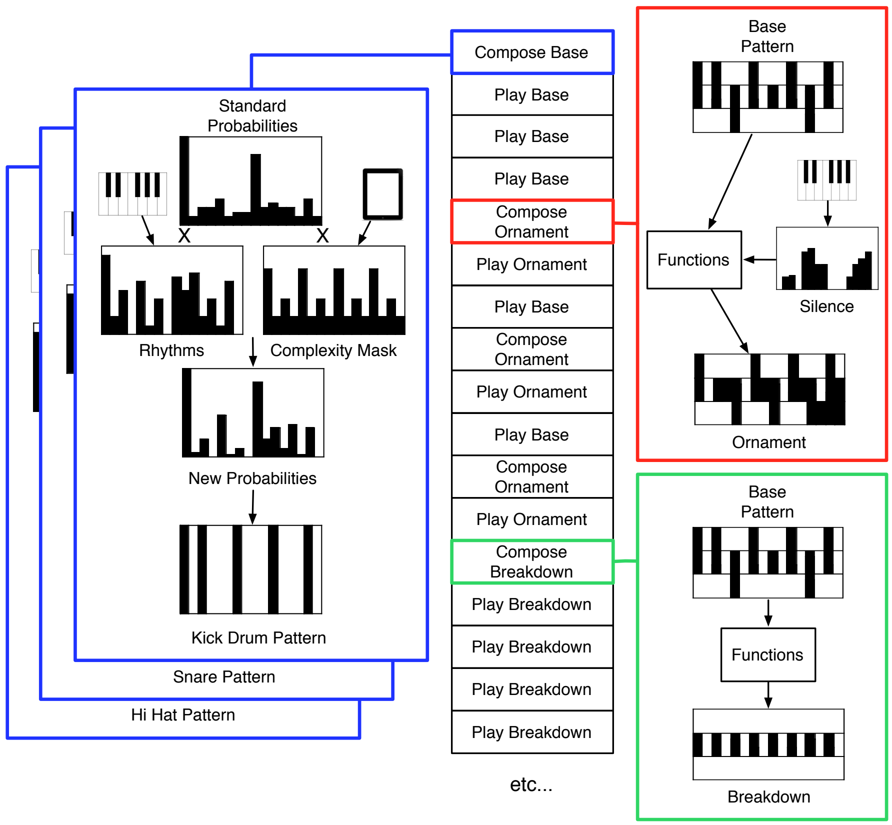

Figure 5.4: Diagram of the Composition of a Single Chorus by Mortimer's Responsive Composition Algorithm

Input Features

At all times in the session a quantised score of participant piano input and robot drum composition is logged. When transcribing the piano input, each note is quantised to the same 16 note grid the robot plays from. Also recorded is the actual timing of the notes, the distance from the beat that it was quantised to, the velocity and the MIDI note value. These are then used to generate some higher level features to influence the composition process. Covered in Section 5.2.2, this includes the composition of base grooves and the choice of ornamentation functions.

$$P(O|t\%4{=}0){=}c{*}3$$
$$P(O|t\%2{=}0){=}(c{-}(1/3)){*}3$$
$$P(O|t\%1{=}0){=}(c{-}(2/3)){*}3$$

Figure 5.5: Equation for Calculating the Probabilty of Mortimer Ornamenting the Base Drum Pattern for a Given Bar (P(0)) Given the Human Inputted Complexity Parameter (c)

Table 5.1: Possible Structures and Probability of Occurence for Chorus Sections Composed by Mortimer

Bars of Verse	Bars of Breakdown	P(x)
12	4	0.5
16	0	0.2
8	8	0.3

5.2.2 Composition

Generating Form

The FormGenerator composes the structure of the piece. This constitutes which sections will be played where and for how long. It also lays out where new music will be composed or where variations to existing music will be conducted. The music itself, that is the score played by the robot, is not composed until these specified points, allowing the composition to take into account the most up to date information about the human's contribution.

Each session is made up of tracks, with each track consisting of a number of choruses, each concluded by a breakdown section. As such, each bar within a chorus will either be the replaying of the base groove as is, an ornamented version of the base groove or a reduced version of the base groove.

The length parameter (l) is used to calculate how many choruses there will be in the track, calculated as $1{+}\lceil(l*8)\rceil$. One of the following structures for each chorus shown in Table 5.1 is then picked, weighted with the given probabilities. Once a overall structure has been picked, the structure of each verse calculated. Each bar in a verse may be either the base groove played as composed or an ornamentation of this. The main factor in deciding the split of base and ornamentation bars is the complexity parameter (c). Shown in Figure 5.5, when c is below 1/3, there is a probability (increasing with c) of an ornament every 4th bar. When c is between 1/3 and 2/3, there is certain to be an ornament every

Table 5.2: Base Probability Tables For for Generation of Kick and Snare Patterns

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Kick	1.0	0.1	0.2	0.2	0.3	0.1	0.15	0.15	0.8	0.2	0.25	0.2	0.2	0.1	0.3	0.1
Snare	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.15	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.15

4th bar and there is a probability (increasing with c) of an ornament every 2nd bar. When c is between 2/3 and 1, there is certain to be an ornament every 2nd bar and there is a probability (increasing with c) of an ornament every bar.

Generating Base Pattern

The foundation of the drummer's playing for each track is a base pattern. This is a one bar section represented at a 16th note level in a binary format. The base pattern also has an accompanying durations track which dictates the groove, that is, individual timing deviations from the quantised grid. Each instruments sequence is run on an individual routine allowing for independent and differing grooves for each.

The base pattern for the snare and kick is generated using 0th order Markov approach. Shown in Table 5.2, each semi-quaver position has a manually ascribed probability and is used to stochastically compose a bar of each. Before generation, the base probability tables for snare and kick are augmented by both a histogram of the previous rhythmic input from the human and the explicitly inputted complexity procedure. The quantised piano transcript for the previous 12 bars is turned into a histogram based on note occurrence. The more notes have appeared in each time step over the given period, the higher the probability of a note for the drum in the corresponding location.

Above 0.5, the complexity parameter increases the probabilities of semiquavers and quavers, below 0.5, these are decreased and the probabilities of notes occurring on crotchet beats increases.

There are 3 possible hi hat patterns, each with a static probability of being chosen.

```
if (bar[bd][8]==1) then
   bar[bd][8]=0;
   bar[bd][9]=1;
   bar[sd][8]=1;
end if
```

Figure 5.6: Example Ornamentation Function Used to Add Variation to a Given Base Drum Pattern

Generating Ornaments

Ornaments are variations on the base groove. Each possible ornamentation is defined by a function which is performed on the base groove, a probability of occurrence and a mask detailing approximately where in the bar it will affect. The latter is included so the algorithm will favour an ornament which occurs in a location in the bar where the robot has predicted the human will leave space. To this end, when generating an ornament, a histogram of empty spaces left in previous bars is matched against the ornaments mask, the closer they align, the more chance of that ornament occurring. An example of an ornamentation function is given in Figure 5.6 and a full list is provided in Appendix B.7.

The ornament functions are implemented in series, that is, in turn, each's probability of occurence is updated in accordance to the silence histogram and a decision is made on whether to augment the base pattern with it or not. This leads to a much wider variance of outcomes than if a single function was chosen using a roulette selection approach. For example, the same base pattern could be ornamented differently by the same ornamentation function, depending on which functions had been preceedingly applied.

Generating Breakdowns

In a similar manner to ornaments, breakdowns are generated by passing the base groove through one of a selection of functions. These include removal of kicks or snares from the base groove, generating a half time version of the base pattern or playing only the first half of a bar. Similarly to ornamentation, multiple breakdown functions may be applied and are implemented upon the base pattern in series.

5.2.3 Adjustments

5.2.4 Adjusting for Power

In addition to composing bars based on a form composed at the beginning, we take note density and mean note velocity as a measure of power for the piano playing. If this drops below certain thresholds then instruments are either dropped or thinned by the drummer in order to match a perceived sparsening of the texture of reduction of dynamics. This function is one commonly seen in commercially available auto accompaniment packages such as Rayzoon's Jamstix.

5.2.5 Adjusting for Groove

This is used to get a feel for the groove of the human input and try to match this with the timing deviations of the robot. To ascertain the groove of the human we measure the distance of the incoming notes from the nearest 1/16 note on the quantised scheduler. When a bar is played by the robot, information from the previous 12 human bars is averaged to provide a set of timing deviations that the robot which, in combination with a default mask, creates a groove matched to human's playing for the robot to use.

5.3 Natural Language Generation

In order to include any language, either verbal, as we may wish to use in face-to-face interactions, or written, as we may wish to use in virtual interactions, *Mortimer* will need to implement some Natural Language Generation (NLG). As the generation and processing of natural language are still vast and ongoing research areas in their own right, it was decided to take a minimal approach. By restricting human input to multiple choice answers, the need to process and natural language is removed. This was implemented by the development of a state machine and accompanying script that the human worked through, providing answers through a tablet running a custom interface built with Hexler's TouchOSC.

Dialogue that varies over time increases engagement [Bickmore et al., 2010]. As there is a clear possibility repeating the script within a session and to maintain interest of multiple sessions, we had to use some techniques from NLG

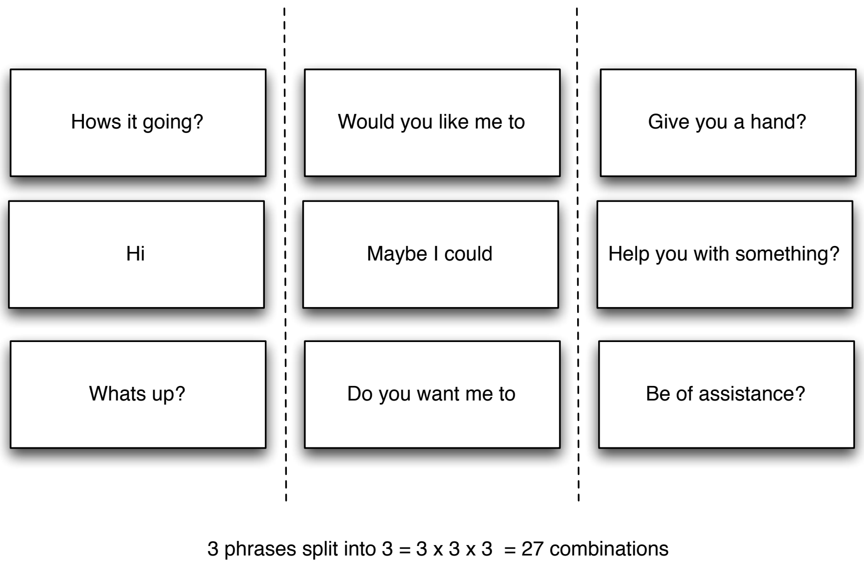

Figure 5.7: Diagrammatic Example of How Phrases Can Be Split into Constituent Parts and Recombined to Increase Variation in Natural Language Generation

to introduce variation into the dialogue. In a simplified version of an approach taken by Skantze and Hjalmarsson, utterances were split into chunks of meaning and a set of variations provided for each chunk. As demonstrated in Figure 5.7, when it comes to recombination, there is now considerably more variation then if we had just picked from a set of complete phrases. A complete dictionary of phrases is provided in Appendix B.6

A speaker is placed in the chest of *Mortimer* and the appropriate text is synthesised using the inbuilt AppleSpeak functionality for Mac OSX.

5.4 Summary

In this chapter we outlined the technical development of *Mortimer*, a responsive robotic drummer, covering the physical construction, composition algorithms and motivations for these choices. In investigation of **RQ2** and **RQ3**, the proceeding three chapters will detail one single session HRI study and two long-term HRI studies in which social behaviours are incrementally added to *Mortimer*. These are evaluated in controlled experiments where humans and

Mortimer improvise music together using the methodology described in Chapter 3.

Chapter 6

Study 1: Framing Human-Robot Musical Improvisation as a Social Interaction

6.1 Introduction

In Section 2.2.1 and Chapter 4 we have shown that musical activity is an excellent candidate for providing an engaging interaction between human and robots. This is in strong support of **RQ2**. **RQ3** seeks to find social behaviours that improve the potential for a positive and sustainable social relationship between human and robot based in music. But do musicians want to have a social interaction with their musical robot? Taking the first steps towards validating **RQ3**, we investigate framing a musical improvisation as a social interaction and hypothesise it will have increased engagement and social presence in comparison to the same system presented with musical functionality alone.

Figure 6.1: Photograph of the Experimental Setup of Study 1

6.2 Method

6.2.1 Participants

Participants were recruited by emailing musical lists and placing adverts on musician recruitment websites. There were 10 participants, 6 male and 4 female between the ages of 26 and 53. 3 classed themselves as beginner pianists, 5 as intermediate and 2 as expert.

6.2.2 Experimental Setup

Two interfaces have been designed for *Mortimer*. One is similar to a tool or instrument where it is controlled through simple stop and play buttons. Another provides a simulated social personality where the sessions are driven by the robot asking questions and providing positive and supportive feedback. Participants were assigned randomly to one of two experimental conditions, A and B. Both groups were briefed that they would be asked to improvise a piece in 4/4 time signature with the robotic drummer. They were told that the study will last 15 minutes but they can leave at any time before that. Subsequently, they

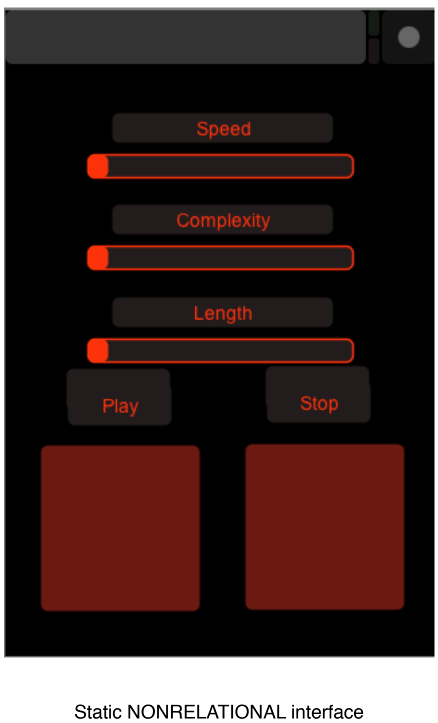

Figure 6.2: Example Screenshots of the Tablet Interfaces Provided to Groups A and B for Study 1

were shown to the room by the researcher where there was a robot and an electric piano, as show in Figure 6.1. A tablet interface is placed on top of the piano, however, as demonstrated in Figure 6.2, the interface and general interaction experience was different for each of the two groups. For Condition B the iPad simply had two buttons labelled "Play" and "Stop", as well as three sliders to control "Speed", "Length" and "Complexity". In Condition A, the robot gave a short autobiographical introduction on arrival in a synthesised male voice. Although this fictional account is technically a lie, research has disproved suggestions that a fictional backstory relayed in the first person may be seen as dishonest [Bickmore et al., 2010].

The robot then invited the human to play and the human could accept using "Yes" or "No" buttons. Reeves and Nass find that people will prefer a flattering computer, even if the praise is unwarranted and prefer a computer which praises others [Reeves and Nass, 1996]. Further, Lee et al. demonstrated verbal behaviour could be used to accurately portray extrovert or introvert personalities and that there was a complimentary personality attraction effect between robot and participants [Lee et al., 2006b]. As such, the following interactions were all framed in a similar way, with the robot providing supportive, positive feedback and politely requesting any changes in behaviour or action. Figure 6.3 shows the flow of the interaction with a complete outline of the state machine available in Appendix B.1. The interface changed appropriately depending on the question and options. The information provided to participants is given in Appendix A.3.

6.2.3 Measures

In Section 3.2.3, we detailed a methodological approach to be taken to evaluate the quality of a human-robot relationships outlining a preferred approach of automated behavioural metrics. We now provide practical examples of measures that can be taken in this specific musical improvisation context. Primarily, we look at the data logs from the robot to examine some quantitative behavioural measures. We also analyse the footage using facial tracking and affect recognition software developed by Soyel and McOwan [Soyel and McOwan, 2013] to ascertain the focus of the participants. In doing so, we hope to observe and measure behavioural changes during the study that display differences in engagement and social presence between the two control groups.

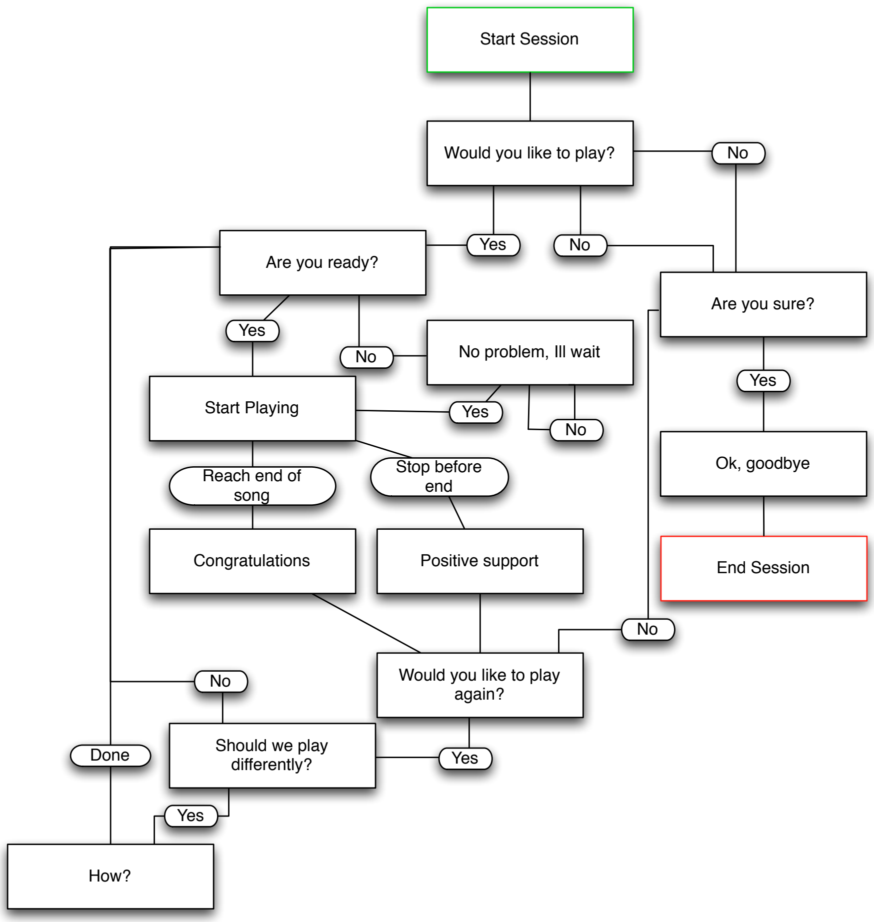

Figure 6.3: Diagram of the State Machine Used by Mortimer During Social Interactions

Although used in later studies, it was not considered appropriate to administer the NRI-SPV survey the because the questions are designed to be responded to about people that they have long-term relationships with. Asking participants questions about someone, whether robot or not, that they have spent 15 minutes with would not be particularly revealing or valid. Below we detail the specific measures taken during the study.

Session Length

We have defined engagement, based upon other research in the HRI field, as the extent to which the human wishes to spend with the robot, as if a person is to develop any type of relationship with a robot then repeated, self motivated and positive encounters are critical. Session length is a fine measure of this, although not without its limitations. It is a broad measure that lacks any real notion of context, making it susceptible to misinterpretations and false positives. For example, Heath tells how the oft used measure of dwell time at museum exhibits favours a long and frustrating experience with a badly designed section over a quick and satisfying one with a clear and concise exhibit [Heath, 2005]. As such, frequency, regularity and length of session should generally by used in tandem with other measures of the quality of interaction.

The length of sessions can either be entirely voluntary with an upper limit, as will be the case in this study, allowing users to leave as soon as they feel, or voluntary past a minimum with an upper limit, as will be case in later studies. Both approaches ensure at least some data in gathered from the session. In the latter case we measure the time that each participant spent with the robot over the minimum required. In our studies, session length is calculated from when the participant first greets the robot via the tablet interface to when the last track of the session ends.

Time Spent Playing

In accordance with **RQ3**, the main focus of our studies is to experimentally and incrementally test the addition of a range of social modalities to supplement the musical improvisation. As such, the split of the sessions between playing and social interaction can be illuminating on a number of fronts. For example, less time spent playing could be indicative of a greater willingness to explore *Mortimer's* social functions and provide evidence that a particular social modality is more engaging than another and in general, more time spent interacting socially would defintitely be an indicator of social presence. Further, as we predict the music will provide a bedrock upon which further facets of a social relationship will develop via the additional social modalities, we would expect the balance of this to shift over time from favouring music to social interactions.

As with most phenomena described here, this measure is open to interpretation and as with others, the combination of this measure in a battery of other additional measures will help remove ambiguity. In this case, less time playing could mean that the musical interaction is not engaging and the user would rather not play at all. Further, it is worth noting that this measure, along with the two that proceed, are limited as although it is something we predict will be an important factor in achieving these ends, an engaging musical interaction in itself does by no mean indicate a successful, positive relationship. As such, in combination with measures of social presence, they can provide evidence to further support the existence of a positive social relationship between human and robot.

Track Endings

Covered in Chapter 5, the tracks within a session may end either by extended breaks in playing, at the natural conclusion of a song, or explicitly by the user via the tablet interface. As such, these can be logged reliably for each user for each session.

Natural stops will demonstrate an increased fluidity of playing, something which could be indicative of the the players entering a flow-like state. Early stops, either by silence or button, will demonstrate that the user has become frustrated or disengaged and would be a signifier of a negative experience. To the detriment of the composition algorithm, it may also be because they have become confused with what the robot is playing. Explicit button stops present a similar interaction one would have with a machine and suggests that the user views the robot as a piece of technology as opposed to a social actor. This would be an indicator of reduced social presence.

Bars Per Track

Similarly to track endings, bars per track is a measure of the fluidity of participant's playing. A higher mean number of bars in a track suggests the user is playing for longer uninterrupted and having a more engaged experience. It also demonstrates that the composition algorithm is working well.

Automated Video Analysis

During the majority of the sessions our participant's hands will be involved in playing music, and not free as they would be in conversation. As they will also be seated throughout and the robot has significantly less degrees of postural freedom than a human, measures of proximity and postural synchrony are unlikely to yield interesting observations. As such, we will be focussing mainly on eye gaze, facial expression and head and body direction.

Several behaviours have been cited as commonly occurring in social situations. This tends to be in the situations where the nonverbal behaviour is used as a socially communicative action. As such, we take occurrences of these as signs of social presence. For example, in a diary study, Provine reports people are 30 times as likely to laugh when with others, 6 times as likely to smile and 4 times as likely to talk [Provine, 1997]. Kraut and Johnston found smiling was more likely to occur amongst bowlers in social interaction than in times of success [Kraut and Johnston, 1979].

The robot has a frontal camera to capture the face and upper body of participants. In order to measure the focus of each participant during the study, we use this footage and Soyel and McOwan's face tracking algorithm based upon Seeing Machines faceAPI [Soyel and McOwan, 2013]. The algorithm can distinguish whether a participant is looking at the robot, the piano or elsewhere in the room. Also, given that context, for example, whether the participant is interacting musically or socially can have a baring on the focus of a participant, we can take each classification and separate them into playing or not playing. The measurement of playing comes from when a track is started until it ends, all other times within the session are considered as not playing.

Manual Video Analysis

As the robot will not have any ASR implemented, it will not respond to any verbal communication, so we do not expect much to come from the human. However, this does mean that if any utterances are directed towards the robot, they may hold even more significance as the participant is fully aware

Table 6.1: Significant Results from T Tests Comparing Quantitative Interaction Data and Automated Video Analysis between Groups A and B for Study 1

Measure	*Mean* A	*Mean* B	$t(8)$	p
Tracks per Session	6.2	9.6	-2.467	0.039*
Bars per Track	58.0	38.5	2.668	0.028*
Natural Stops (%)	87.0	35.6	2.954	0.018*
Button Stops (%)	12.8	61.0	-2.426	0.041*
Focus on Piano (%)	11.1	36.4	-3.928	0.004**
Focus on Piano - playing (%)	10.2	34.4	-4.300	0.003**
Focus on Robot - not playing (%)	47.1	29.3	2.510	0.041*

* p<.05 ** p<.005

that it will have no effect. To record these the sessions were manually coded afterwards.

6.3 Results

6.3.1 Quantitative Interaction Data

All study participants in the B condition stayed for the full 15 minutes, the same is true for Condition A apart from 1 participant who stayed for 13 minutes and 10 seconds. As such we are unable to find any significant difference between the two groups on total length of session. Within the session, we recorded the number of tracks played, defined as a start and stop in play, either triggered manually, or occurring naturally as the robot had reached the end of its pre-composed track. In Condition A a stop could also occur through prolonged silence from the human, in which case the robot would stop and give supportive feedback.

We found significantly less individual tracks within a session for Condition A, $t(8)$=-2.467, p=0.039, as well as significantly more bars per track in Condition A, $t(8)$=2.668, p=0.028. Further, we found that there were significantly more natural stops in Condition A $t(8)$=2.954, p=0.018, as well as significantly more manually triggered stops in Condition B $t(8)$=-2.426, p=0.041.

We did not find any significant differences in the mean values of any the performance parameters or the frequency with which they were altered, implying

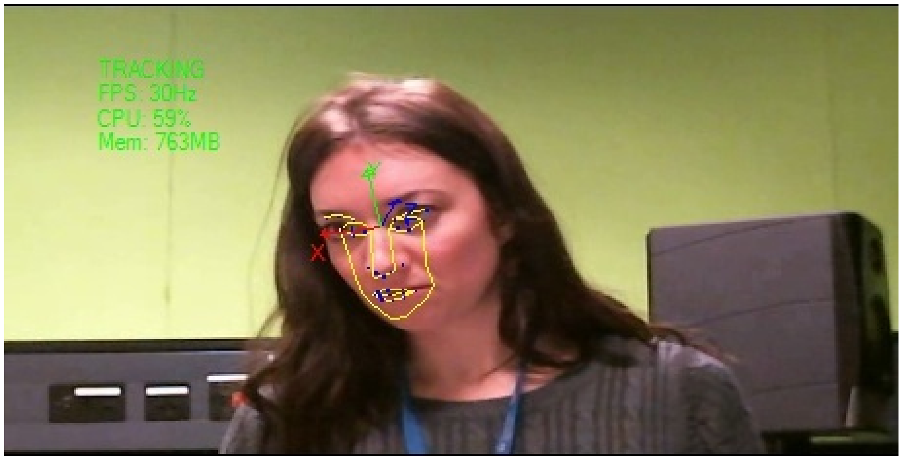

Figure 6.4: Screenshot of McOwan And Soyel's Face Tracking Being Used on a Video of a Participant From Study 1

that the experimental condition did not have any bearing on how the participants explored them.

6.3.2 Automated Video Analysis

As demonstrated in Figure 6.4, we analysed all footage from the frontal camera using Soyel and McOwan's affect recognition software [Soyel and McOwan, 2013]. Based on Seeing Machine's faceAPI, it provides us with data as to the focus of the participant's attention and also infers an affective state. With regards to focus, it distinguishes between the piano, the robot and elsewhere. Initial tests for differences between the groups based on mean times spent in each state across the whole session indicated that Condition B spent significantly more time looking at the piano, $t(8)$=-3.928, p=0.004, and spent significantly more time in a thinking state $t(8)$=3.402, p=0.009. All other results showed no significant difference between groups. However, social and task-related context is crucial when attempting to infer information about the user's affective state or level engagement [Castellano et al., 2012] so it is arguable how much insight we can gain from a contextless measure that completely ignores what else is happening during the session. For instance, looking at the piano while playing could suggest concentration, whilst looking at the piano while not playing could suggest shyness or avoidance. Or, since the interface was placed on top of the piano, spending time looking at this when not playing could be explained by participants changing parameters.

To investigate further, we measured the mean time spent in each state and took into consideration whether music was being played or not. Any change in behaviour during playing is of particular interest as both groups were interacting with exactly the same system once the music began. Once the conditions had been separated, we found that the difference in attention towards the piano was only significant when the participants were playing, $t(8)$=-4.300, p=0.003. This is perhaps explained by the user thinking about the robot as nothing more than a sound making device, as one would a loudspeaker, and becoming more immersed in their own playing and the tablet. Further, we found that Condition A spent significantly more time looking at the robot when they were not playing $t(8)$=2.510, p=0.036.

With regards to the affective states, the results showed significantly more time in the thinking state when playing for Condition B $t(8)$=-3.758, p=0.006 and significantly more aggression when not playing $t(8)$=3.011, p=0.017.

6.3.3 Manual Video Analysis

When coding for social presence in a multi-session HRI trial, Leite et al. recorded verbalisations both in response to questions and when uninitiated by the robot [Leite et al., 2009]. We have also noticed verbalisations during the sessions from both groups and manually coded for these, taking it as a signifier of social presence in a system where the user knows the robot cannot hear or understand any verbal communication. 2 participants in Condition A consistently engaged in back chatter (8 and 19 times) in response to questions. 2 participants from Condition B and 1 from Condition A also greeted the robot verbally before playing. Analysis was completed by three coders and inter-rater reliability was assessed using a two-way, mixed, consistency, average-measures ICC. The resulting ICC was in the excellent range, $ICC = 0.998$, indicating that coders had a high degree of agreement.

6.3.4 Validation of Automated Video Analysis

We have used Soyel and McOwan's algorithm for our automated video analysis, however, it was originally trained for detecting affect in children playing chess with a robot. We expected it to be robust enough to fit our similar context however, a validation study with human coders was conducted to confirm this. A dataset of 90 still frames was collated from the participants of Study 3. This study is described in Chapter 8 and used the same experimental setup and video analysis as this study and Study 2 (described in Chapter 7). It was chosen as a source dataset for validation as it was a long term trial with the largest number of participants so provided the greatest variety. The frames were selected by randomly picking a week for each class for each participant. For this week the 3 longest uninterrupted sequences of the class were found and the midpoint chosen as an example frame for that class. This results in 3 frames of each class for 10 participants and a dataset of 90 images.

Coders, all PhD candidates from the Cognitive Science Research Group at Queen Mary University of London, were asked to classify each images into one of the 3 following categories.

1. Head and eyes down and straight ahead
2. Head and eyes straight ahead
3. Neither of the above

Class 1 is interpreted as the piano, Class 2 as the robot and Class 3 as elsewhere. The coders were also given 6 example frames (not included in the dataset) for each class. The information sheet provided is included in Appendix A.5.

The mean classification success was 88%. Again, inter-rater reliability was assessed using a two-way, mixed, consistency, average-measures ICC. The resulting ICC was in the excellent range, $ICC = 0.92$, indicating that coders had a high degree of agreement. This demonstrates that the algorithm was satisfactorily transferrable to our context and able to provide the measures desired.

6.4 Discussion

Section 3.1.3 defined a human-robot relationship as the development and maintenance of social presence and engagement over multiple interactions where the human party displays some behaviours indicative of a positive interpersonal relationship. As this is only a single session study we cannot comment on the maintenance of these factors, however, it is this model we will use when analysing how the results in Section 6.3 can provide answers to **RQ2** and **RQ3**.

As nearly all the participants completed the full session, this shows that regardless of social framing, the musical interaction was at least temporarily engaging, suggesting an affirmative answer to **RQ3**. Beyond this, having found that with the social framing there were less tracks that lasted for more bars, we suggest that Condition A played with greater fluency and were more engaged. This is enforced by the greater number of natural stops and lower number of button stops. These latter findings also demonstrate a greater social presence as the participants treated the robot less like a machine or instrument and more like a social actor, waiting for the natural pause rather than stopping mid-performance.

Findings in Section 6.3.2 that those in Condition A looked more at the robot when not playing is perhaps to be expected as the robot was drawing attention to itself by speaking to the user. However, we evidenced in Section 3.2.2 that looking at a robot has shown to be an accurate metric for engagement in previous HRI trials and so we take it as such.

The apparent significance of playing and not playing highlights the importance of the distinct roles of mutual gaze in musical and conversational interaction. Gratier says that although gaze is necessary for 'grounding' in conversation, improvising musicians do not need to see each other for this purpose [Gratier, 2008]. As such, our finding that focussing on the robot was only a significant difference when not playing is unsurprising and does not detract from our inferences as to the effect of the experimental condition on engagement and social presence. Additionally, our finding that those in Condition A spent more time looking at the robot when not playing is also an expected outcome due to the importance of gaze in managing social interactions and is indicative of participants behaving as they would socially with another human. We interpret this as evidence of feelings of social presence towards the robot when presented as a social actor.

We do not put a huge amount of weight in the findings with regards to affective state as the definitions of these come from the context of children playing chess and so are not necessarily transferrable to our domain. This being said, Thompson et al. suggest that facial expressions are often used as *affect displays* in musical performance [Thompson and Graham, 2005], meaning that automatic facial recognition techniques could be used if developed for a musical context.

Although their response is initiated by the robot, results in Section 6.3.3 are interpreted as a sign of social presence as they were not told the robot could understand their speech, nor did any its behaviour indicate this. The verbal greeting is perhaps a more interesting finding as those in Condition B were not expecting any interaction at all and so the social behaviour is triggered purely by the anthropomorphic form of the robot

6.5 Conclusion

We have shown that presenting an IMS as a social actor rather than as a instrument changes the way people play and behave. We have found they play with more fluency and are less likely to stop and start the robot mid performance in the former condition. They also look at the robot more when not playing and look at the piano less when playing. We suggest that these results show greater engagement with the robot and the playing and greater sense of social presence when presented as even a rudimentary social actor. Even though the study was a single, short session and so susceptible to the novelty effect, the results are promising in the context of **RQ3** and embolden us continue along this path, extending to long-term studies with additional simulated social behaviours.

Chapter 7

Study 2: How Nonverbal Behaviour Improves Human-Robot Relationships

7.1 Introduction

Results from Chapter 6 showing framing a session between pianists and our drumming robot *Mortimer* as a social interaction resulted in greater feelings of engagement and social presence. Although these results are encouraging, given the model of a human-robot relationship presented in Section 3.1.3, both **RQ2** and **R3** require the examination of these factors over multiple interactions. As such, we present a similar study, extended to long-term trial with 6 weekly sessions for each participant.

In further investigation of **RQ3**, we also plan to add more simulated social behaviours to *Mortimer*. In their canonical survey of the field, Fong et al. cited realistic facial expression as a key design factor in social robots, especially in the demonstration of affective behaviour [Fong et al., 2003]. Further, being able to communicate and interpret nonverbal actions can be crucial to the success of social interactions [Guerrero and Floyd, 2006]. Noller extends this by claiming nonverbal communication is important for maintaining social bonds, as it allows people to express emotions and to relay how they feel about each other and the relationship [Noller, 2006]. She also reports that the accuracy of decoding of nonverbal cues is often a predictor of relationship closeness and satisfaction. Tickle-Degnen suggests that nonverbal expressivity on the whole tends to have positive social outcomes, including rapport [Tickle-Degnen, 2006]. Within this, Fridlund and Russel claim that faces play a key part in our social interactions [Fridlund and Russell, 2006], indeed, interpreting and imitating facial expressions is one of the first skills an infant learns [Rinn, 1991]. Weinberg, Raman and Mallikarjuna describe a social and musical interaction between human musicians and two robots, one with virtually embodied head movements, however, this is only one session and no evaluation is presented [Weinberg et al., 2009]. Motivated by this, a set of head poses and facial expressions triggered by social and musical cues was developed for *Mortimer*.

6 sessions per participant were conducted in order to study the effect of the experimental condition on, and the suitability of musical improvisation in general for, developing and maintaining a positive human-robot relationship. Following the methodology developed in Chapter 3, automated behavioural metrics were mainly used, analysing data logs to see how participants interacted with the robot and using face tracking to determine where they are focussing their attention during the sessions. In relation to a control group, we expected the inclusion of head poses and facial expressions to increase social presence and engagement within the sessions, seen by increased session time, smoother playing and more displays of behaviour indicative of an interpersonal relationship.

7.2 Related Work

Before describing the implementation of musically and socially triggered facial expressions and head poses in *Mortimer*, we provide evidence as to the important role they play in both contexts.

7.2.1 Nonverbal Cues in Musical Performance

Nonverbal cues, notably facial expressions, mutual gaze and head movements are used by musicians to convey information about the music either to co-performers or audience members [Cicconet et al., 2013]. This plays an especially important role in improvised music.

In almost all acoustic music performance, the body, and in some cases the head and face, are inseparably coupled to the generation of sound [Vines et al., 2006, Thompson and Graham, 2005, Gratier, 2008]. However, they are also used as cues, intentionally or not, to augment the performance and to anticipate or accentuate important events. For example, in an analysis of an improvising jazz guitarist, Gratier demonstrates that musicians may use their body movements to convey the structure and meaning of the music [Gratier, 2008]. Similarly, Vines et al. discovered that the perceived tension of a performance is most influenced by visual, rather than auditory, cues [Vines et al., 2006]. They also report that it is a combination of auditory and visual stimulus that affects audience's perception of phrasing in a musical performance, providing the supporting observation that the contours of the performer's body movement tended to align with their phrasing of the music. Further, Thompson et al. find that facial expressions are used to convey timing events, thus increasing musical intelligibility [Thompson and Graham, 2005]. They also report that facial expressions can be used to make music sound more or less dissonant or to make musical intervals sound further apart or closer together.

Gratier suggests that facial displays of affect may serve the purpose of grounding between improvisers. For example, a musician may smile at a mistake or a particularly satisfying lick [Gratier, 2008]. Moreover, whilst drawing comparisons between improvised music and conversation, she reports that mutual gaze is much less constant in the former. This being said, although less frequent, it still serves a crucial role in managing the interaction and tends to occur during moments of structural change or importance in the music.

In a study of a performance by blues guitarist BB King, Thompson et al. find he often used facial expressions to display affect. For example, in moments of tension he takes on an introspective demeanour, looking down and shaking his head. A musicologist interprets this as him signalling he feels the emotion but will not submit to it. Alternatively, in moments of release he opens his mouth towards the audience as if in wonder. As well as relating to affect, they find King's head movements often react to individual notes and licks and tend to reflect only his performance, rather than that of his band. A study of a Judy Garland performance by the same authors reveal how she uses hand gestures in a more illustrative fashion, literally reflecting the lyrics of the song, displaying the range of purposes bodily movement can play for different performers.

7.2.2 Facial Expressions in Social Interaction

Since the early 1960s, psychologists have prevalently viewed the face as the key factor in understanding the emotions of humans. However, Chovil makes the argument that facial expressions are not primarily, or even at all, expressions of an internal affective state but serve the purpose of being socially communicative actions [Chovil, 1997]. Kraut and Johnston demonstrated that smiles were more likely to occur during social interaction than in situations of happiness in a study of ten-pin bowlers [Kraut and Johnston, 1979]. Further, analysing gold medal ceremonies, Fernandez-Dols and Ruiz-Belda found that a greater proportion of smiles occurred in the interactive stage of the event than elsewhere. This is surprising, considering the whole event is assumed to be one where the athletes will feel intense joy throughout [Fernández-Dols, 1997]. The rejection of the emotional cause for facial expressions is taken the extreme by Fridlund and Russel, who introduce the Behavioral Ecology View (BEV) [Fridlund and Russell, 2006], providing an alternative socially communicative explanation for all the expressions which others have claimed are ”readouts” of prototypical emotions. For instance, smile moves from ”readout of happiness” to a signifier of ”readiness to affiliate or play”[1] and ”readout of anger” becomes the message ”readiness to attack”[2]. Under Fridlund and Russel's treatise, *Mortimer* should use his face to reflect planned intentions and goal states, not emotions.

Regardless of the intention, be it internal affective mirror or socially communicative gesture, it is worth examining what information a face can reliably relay to others within a social interaction. It is reasonable to suggest the face can allow us to distinguish between pleasant and unpleasant expressions and between differing degrees of these expressions [Ekman and O'Sullivan, 1991]. Beyond this, there is good evidence to show that at least 6 distinct facial expressions can be universally distinguished and recognised [Ekman and O'Sullivan, 1991] and these have been classed as happiness, sadness, surprise, disgust, anger and fear. Smith and Scott outline a further componential model which defines 6 types of behaviours and how they can be expressed [Smith and Fernández-Dols, 1997]. This includes pleasantness, goal obstruction, anticipated effort, attentional activity, certainty, novelty and personal agency and draws from not only their own research but various historical models.

[1] [Fridlund and Russell, 2006]
[2] [Fridlund and Russell, 2006]

Fernandez-Dols and Carrol demonstrate that although much research treats it as such, it is inherently problematic and reductive to consider facial expressions outside of their context [Fernández-Dols and Carrol, 1997]. If we are to clearly and unambiguously use the face of the robot to demonstrate social and musical cues and emotions then we must be aware of the context that they are being produced in, otherwise they may fail to be interpreted as intended. Luckily, in our laboratory experiments, the context is known and controlled to a high degree.

Human's ability to decode nonverbal behaviour can be further influenced by their knowledge of display rules of the specific person, helping them to judge intensity or simulation, and identify posed or ironic displays. Individual differences can also come from gender, race and age [Burgoon and Bacue, 2003].

7.2.3 Head Movements in Social Interaction

Head movements are far from arbitrary, they have been shown to reliably occur at certain points in social interactions, serving many functions from emblematic replacements of speech, to turn management and backchannelled affirmation [McClave, 2000]. As a general rule, speaker's heads tend to be in constant motion whereas listeners tend to be relatively static.

In a microanalysis of a corpus of filmed social interactions, McClave found several consistent co-occurrences of head movements and social cues [McClave, 2000]. For example, a lateral sweep is used to demonstrate inclusivity, often concurrently with words such as "everyone" and "whole". Repeated head movements also often coincide with listed items when a speaker is delivering alternatives. Further, head shakes, as well as serving the emblematic purpose of negation, are often used during speech to emphasise a sentiment more intensely or to express uncertainty. This was also seen by Iwano et al., who found that horizontal head movements occurred during denials [Iwano et al., 1996]. Similarly, both found that a head nod, or vertical movement, is often used to demonstrate affirmation, agreement and continuing comprehension. Iwano et al. also found that when speakers are expecting a response, such as preceding a question, they often lift their head up to face their partner directly [Iwano et al., 1996]. In terms of persuasiveness, Briol and Petty find that nodding and head shaking during conversation stands to strengthen or undermine your argument respectively [Briñol and Petty, 2003]. Head movements can also provide attentional

Figure 7.1: Photograph of Mortimer with Head Updated for Study 2

Table 7.1: Description of and Triggers for Face and Head Movements in Mortimer

What	When	Why
Smile	When you have answered a question with a positive outcome	Smiles used as backchannels
Smile	When positive reassurance is being offered	Agreement
Smile	Following a breakdown	Release
Raised Eyebrows	Before a question	Shows inquisitiveness
Closed Eyes, Eyebrow Frown, Tight Mouth	During breakdown	Shows Tension
Closed Eyes, Eyebrows Raised, Smile	During breakdown	Shows Transportation
Eyebrow Frown	Complicated Ornament	Shows Concentration
Head Nod	When you have answered a question with a positive outcome	Shows Agreement and Affirmation
Head Leans Back	During breakdown	Shows Transportation
Move Head to Side To Side	Complicated Ornament or Breakdown	Shows Intensity
Lean Forward	After question	Demonstrates response expected

cues that make up our sense of engagement with another [Michalowski et al., 2006].

7.3 Implementation

7.3.1 Facial Expressions

LaFrance suggests that the causes of facial expressions are far more complicated than the usual ”readout” approach that most computer scientists take [LaFrance, 2008] and the lack of a clear and consistent link between an internal emotional model and facial expressions leads us to approach any such system with caution. However, we have shown in Section 7.2.2 that facial expressions can be used with satisfactory accuracy and universality to broadly express negative or positive emotions, as well as other more practical social cues such as attention and interestedness.

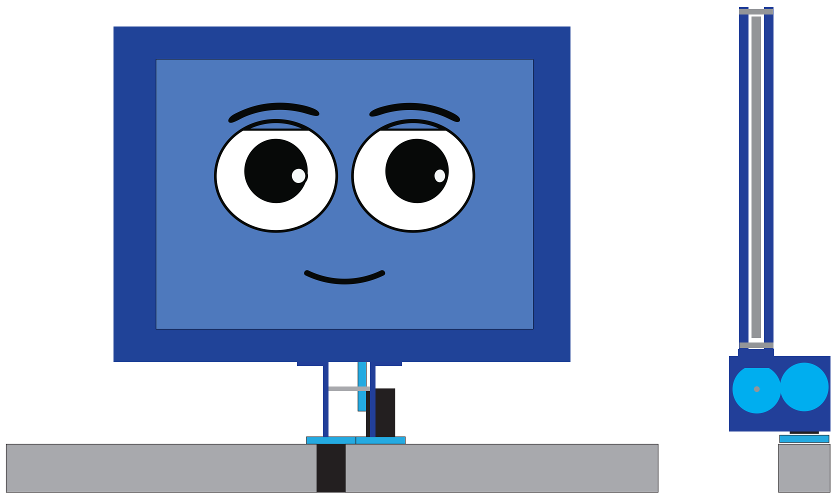

Figure 7.2: Technical Drawing of Mortimer's Pan-Tilt Head Mechanism

Following findings in Section 7.2.1, *Mortimer's* face was used to reflect moments of tension and release in music, as well as moments of concentration. These expressions were also used during musical performance to aid mutual comprehension as the robot enters and exits breakdown sections.

In terms of technical implementation, Fong et al. report that this is often not done well and describe mechanical approaches as often clunky and abrupt [Fong et al., 2003]. Further, Delaunay et al. suggest the mechanical complexity often comes at a great cost in development and maintenance [Delaunay et al., 2009], also, that mechanical android faces are yet to reach levels of humanness necessary to avoid the uncanniness that can lead to anxiety and unease. In fact, this is something to be wary of when attempting any humanoid face, even with smoother animated approaches, such as Brennand and Gordon's Mask Bot [Brennand and Gordon, 2012]. This being said, using the mechanically faced EMYS robot, Ribeiro and Paiva managed to get high classification rates for 5 out of 6 emotions inspired by Ekman's descriptions of distinguishable facial expressions [Ribeiro and Paiva, 2012].

Given the importance of context and the negative effects of misclassification, we aimed to design facial expressions that are clear and unambiguous in what they attempt to convey and that they occur at appropriate times in concordance with other appropriate actions. As such, a small screen was used to allow complex realtime animations that are smooth and easily changeable. To avoid the afore mentioned uneasiness associated with the uncanny valley [Delaunay et al., 2009], a simple, cartoonish face was chosen using the basic facets of, but clearly not attempting to replicate, a human face. The eyebrows and the mouth are focussed on as amongst the most reliable and regularly used in facial expressions. For example, there us strong agreement between researchers that an eyebrow frown is a sign of negativity or concentration and a smile is a signifier of pleasantness [Smith and Fernández-Dols, 1997]. Further, those who show more positive expressions of affection are more likely to be rated as having good nonverbal skills [Guerrero and Floyd, 2006, Noller, 2006] so positive facial expressions such as smiles were favoured. Animating the mouth also serves a practical purpose for dialogue.

The facial expressions *Mortimer* uses and their triggers are detailed in Table 7.1 and Figure 7.3 with a complete outline in Appendices B.1, B.2, B.5 and B.4.

7.3.2 Head Movements

As well as looking to human's use of head movements to influence our design, we are also instructed by previous work in robotics. For instance, Macdorman and Cowley demonstrated that attentive head movements are sufficient to elicit the perception of what they call personhood, a concept that we have shown to have large overlaps with social presence and believability [Macdorman and Cowley, 2006]. Head movements have also been used by Weinberg et al. in their musical robot *Shimon* in order to increase its social presence within an ensemble [Weinberg et al., 2009]. Similarly, Szafir and Mutlu report increased engagement when adaptive head movements are included in a teacher student HRI trial [Szafir and Mutlu, 2012]. Breazeal and Fitzpatrick use leaning forwards or recoiling back with the head in order to show willingness to engage or fear, allowing their robot *Kismet* to regulate its personal space [Breazeal and Fitzpatrick, 2000].

Mounted on a pan/tilt device constructed from two servo motors, *Mortimer's* head has two degrees of freedom. The head movements *Mortimer* uses and their triggers are detailed in Table 7.1 and Figure 7.3.

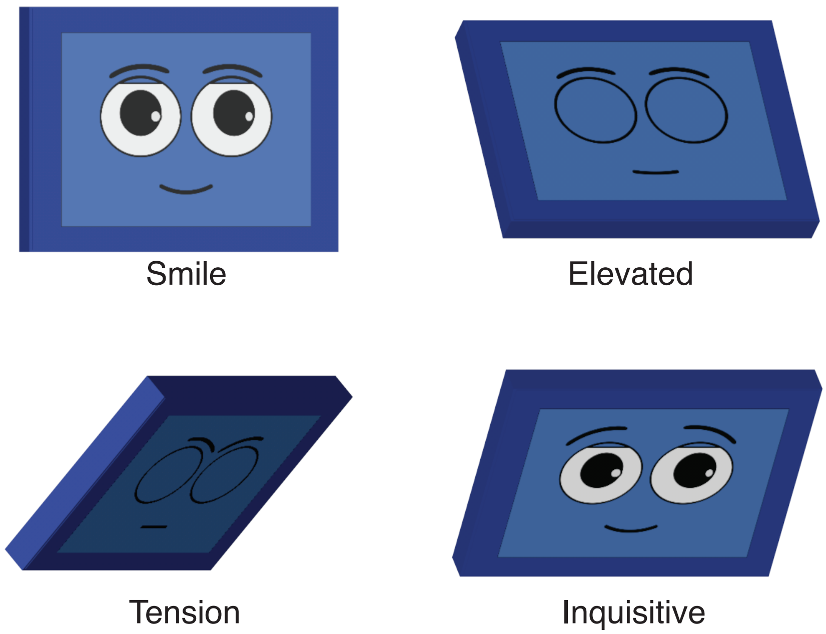

Figure 7.3: Renderings of Selected Robotic Facial Expressions and Head Poses Used by Mortimer in Studies 2 and 3

7.4 Method

7.4.1 Participants

Participants were recruited by emailing musical lists and placing adverts on musician recruitment websites. There were 10 participants, 5 male and 5 female between the ages of 22 and 54. There was a wide range of self reported skill level (1-5=beginner-expert, min=1, max=5, mean=3.1, SD=1.29). Even though the number of participants is relatively small, due to a practical constraint of needing skilled participants, as each returned multiple times we conducted 60 sessions in all.

7.4.2 Experimental Setup

Participants were asked to attend 6 identical weekly sessions. After an initial 30 minute session, at each proceeding session they were informed they had to stay for a minimum of 20 minutes, after which they might leave and still fulfil the study requirements. They could also continue to play for anything up to another 25 minutes, leaving at any point. Participants were recompensed £50 upon completion of the study.

During the sessions, participants could freely improvise with *Mortimer*, who facilitated the interactions with a rudimentary artificial personality. The participants were randomly assigned to one of two experimental conditions. For those in Condition C, the robot included all the head movements and facial expressions detailed in Section 7.3, whilst for those in Condition D, the head and face remained static throughout. The information provided to participants is given in Appendix A.3.

7.4.3 Measures

Automated Behavioural Metrics

Following the methodological approach outlined in Section 3.2.3, a multitude of quantitative interaction data was recorded during the study. This included the time that each participant spent with the robot over the minimum required 20 minutes, the number of button stops and the proportion of the session the participant would spend interacting musically or socially.

We will also record the number of interruptions during the session. During the social facilitation of the sessions, the questions posed by *Mortimer* are both verbalised and displayed on the user's tablet. Both methods were chosen to combine the clarity of a written instruction with the increased engagement of varied vocal dialogue [Bickmore et al., 2010]. As soon as the screen for this particular stage appears the user may answer the question on the tablet, interrupting the speech before its conclusion. This may occur when participants are able to read the question in less time than it takes for *Mortimer* to vocalise and will be exacerbated within and across sessions as they become increasingly accustomed to sequences through repeat interactions. An interruption is recorded if a question is answered before *Mortimer* has spoken every word of the planned speech. A limitation of this measure is it can only be used if the experimental condition involves the social interaction.

Interruptions of vocalisations will show that participant is happy to interact with the robot through a touch screen interface as they would a machine, rather than through, albeit onesided, dialogue as they would in a social interaction.

This would not necessarily be to the detriment of inferred engagement but almost certainly to suggestions of social presence.

Unlike our previous study, where all but one of the sessions were of equal length, the length of sessions ranged from the minimum of 20 minutes right up towards the maximum of 45. As such, the measure of tracks per session used previously is confounded by this variable and not a particularly elucidating one when attempting to investigate the smoothness and immersion of the participant's playing. However, the measure of mean bars per track provides us with a measure of average length of tracks within a session independent of session length.

In order to measure the focus of each participant during the study, we again used Soyel and McOwan's face tracking algorithm based upon Seeing Machines faceAPI [Soyel and McOwan, 2013].

Self Report

During our survey of questionnaires used for evaluating human-human relationships, the NRI-SPV is the one we think is the one with the best potential for use in human-robot relationship trials. Although we are generally unsupportive of questionnaire-based approaches, the NRI has been validated thoroughly alongside behavioural studies. Further, a human-robot relationship is necessarily a one-sided affair so we are only interested in the human's perceptions of their relationship with a robot. The NRI-SPV survey gives us this with a detailed breakdown of a number of perceived provisions.

As it is reflective in nature and addresses quite broad aspects of a relationship, it is not appropriate for tracking week on week changes. However, administered at the midpoint and on completion of the trials, it may provide some insight into which provisions are perceived as being given by the robot to participant at these points and if there are any differences between experimental groups. We would expect the highest rated provisions to be those previously discovered to be imparted in close relationships and music, namely, reassurance of worth (WOR) and instrumental aid (AID).

In Section 3.2.2, we cited that Bickmore et al. claim self reported intention to repeat interact can be a useful measure of how well a socially interactive system

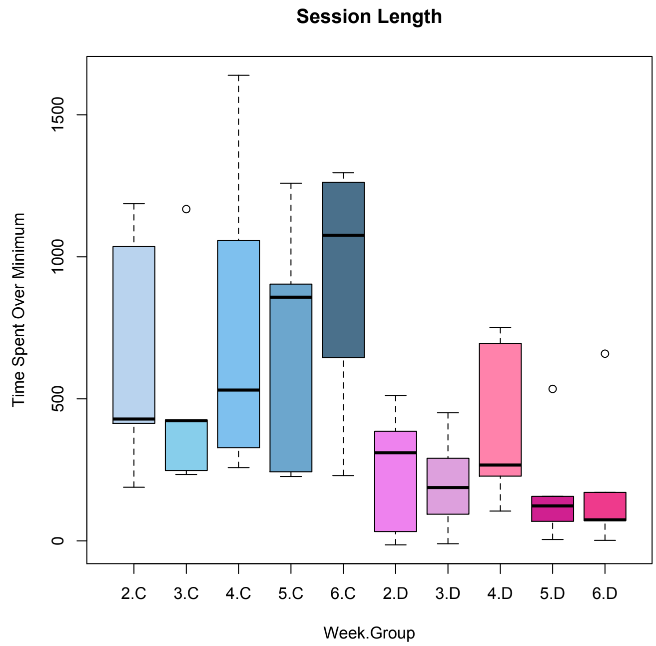

Figure 7.4: Session Length For Groups C and D in Study 2

is suited for long-term engagement [Bickmore et al., 2010]. The question is posed to participants as part of an exit questionnaire.

7.5 Results

7.5.1 Quantitative Interaction Data

To measure the effect of experimental condition on the data gleaned from the data logs and its change over time we fitted a random intercept linear mixed effect model for the fixed effects of week, group and the interaction of the two. Results are displayed in Table 7.2.

We found significant effect of group ($\beta = -456.52$, 95% CI [-751.91 -163.77], p=0.047), demonstrating that those in Condition C voluntarily spent more time with the robot. The interaction of group and week was also significant ($\beta = -82.46$, 95% CI [-155.23 -14.06],p=0.035), demonstrating that the way that those in Condition C changed the amount of time they spent with the robot over the study period differed positively from those in Condition D. For the group

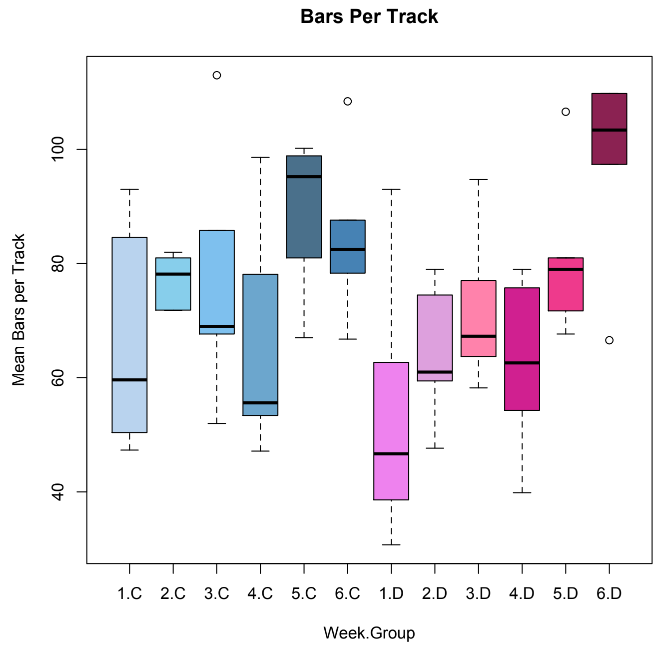

Figure 7.5: Mean Bars Per Track For Groups C and D in Study 2

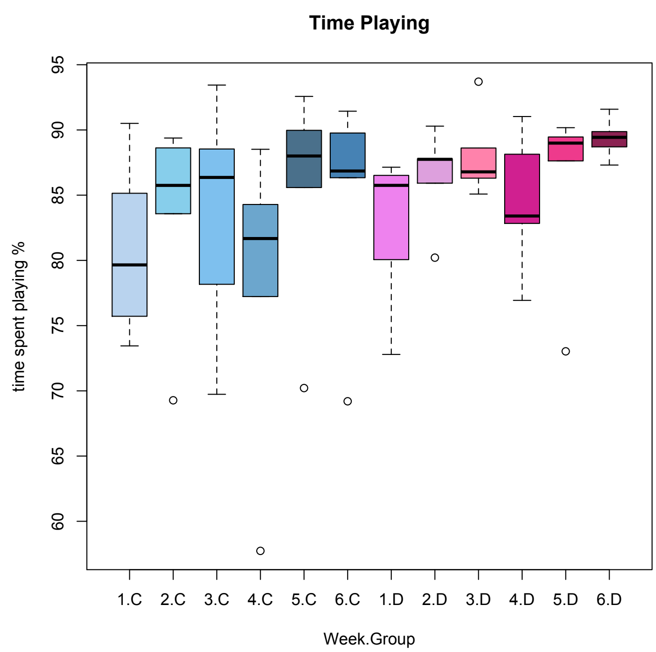

Figure 7.6: Time Spent Playing For Groups C and D in Study 2

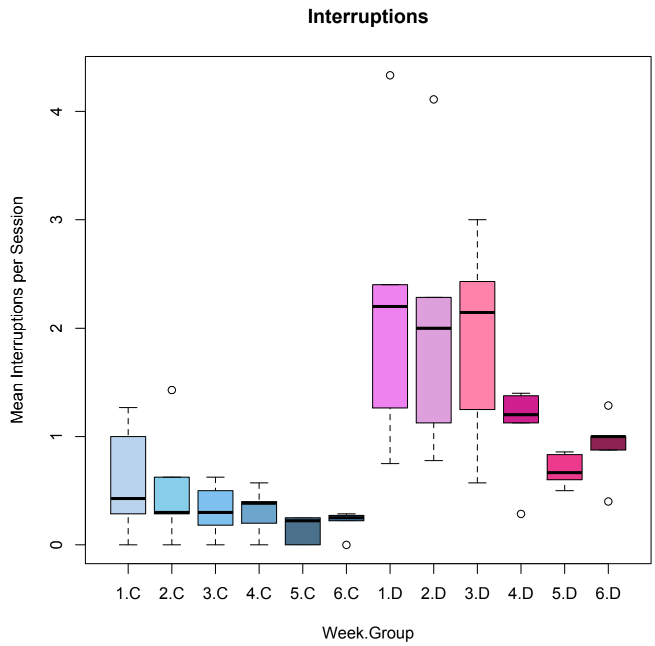

Figure 7.7: Interruptions For Groups C and D in Study 2

as a whole the mean number of bars per track increased over time, meaning longer tracks and less interruptions during playing (β = 5.26, 95% CI [3.67 6.93],p=0.0005). Interestingly, the rate of increase was greater for those in Condition D (β = 4.11, 95% CI [0.9 7.24],p=0.0005). With respect to the proportion of time session spent playing piano, we found a significant effect of week (β = 0.7, 95% CI [0.19 1.19], p=0.0265), demonstrating that regardless of the experimental condition, all participants spent less time playing with the robot as the study progressed.

For interruptions, we found a significant effect of group (β = 1.12, 95% CI [0.69 1.56], p=0.0045), demonstrating those in Condition C interrupted the robot less over the whole study. There was also a significant decrease in number of interruptions across the trials ($\beta = -0.21$, 95% CI [-0.28 -0.14], p=0.005). Further, the rate of reduction of interruptions over the trial was significant higher for those in Condition D ($\beta = -0.23$, 95% CI [-0.37 -0.11], p=0.005). However, the proportion of button stops presented no significant effects for week, the experimental condition or the interaction between the two.

Table 7.2: Results of Random Intercept Linear Mixed Effect Model For Quantitative Interaction Data From Groups C and D in Study 2

Data	**Fixed Effect**	**Estimate β**	**CI [5% 95%]**	p
Session	Week	28.77	[-7.29 66.78]	0.2184
Length	Group	-456.52	[-751.91 -163.77]	0.047*
	Week.Group	-82.46	[-155.23 -14.06]	0.035*
Bars Per	Week	5.26	[3.67 6.93]	0.0005***
Track	Group	-4.91	[-17.09 8.01]	0.5832
	Week.Group	4.11	[0.9 7.24]	0.0005***
Button	Week	-0.01	[-0.02 0.01]	0.3913
Stops	Group	0.06	[-0.01 0.12]	0.2404
	Week.Group	-0.02	[-0.05 0]	0.2649
Inter	Week	-0.21	[-0.28 -0.14]	0.0005***
-ruptions	Group	1.12	[0.69 1.56]	0.0045***
	Week.Group	-0.23	[-0.37 -0.11]	0.0005***
Time	Week	0.7	[0.19 1.19]	0.0265*
Playing	Group	3.55	[-2.2 9.34]	0.4158
(%)	Week.Group	0.28	[-0.73 1.26]	0.1579

Random Intercept Linear Mixed Effect Model for quantitative interactional data. P values are estimated from a parametric bootstrap (2000 replicates). Confidence Intervals are estimated from a parametric bootstrap (2000 replicates). *p<0.05, **p<0.01, ***p<0.001.

7.5.2 Automatic Video Analysis

We fitted a random intercept linear mixed effect model for the fixed effects of week, group and the interaction between the two for each category. Results are displayed in Table 7.3.

We found significant effect of group ($\beta = 28.75$, 95% CI [12.72 45.56], p=0.036) and the interaction between week and group ($\beta = 4.11$, 95% CI [1.05 7.53], p=0.015) for proportion of time spent looking at the robot when playing. There was also an effect for group for looking at the robot when not playing ($\beta = 26.28$, 95% CI [14.91 38.71], p=0.0125). This demonstrates that over the course of the whole study, those in Condition C spent less time looking at the robot whether playing or not. Also, that way the two groups differed in the former category changed as the study progressed.

The only other significant effect was for week for looking at the piano when not playing, ($\beta = 1.259$, 95% CI [0.44 2.15], p=0.0235), with those in Condition C looking at the piano more when not playing.

7.5.3 Self Report

We conducted T-tests for each provision and the amalgamated positive and negative scores of the NRI-SPV. Analysis of results did not find any significant factors between groups or over time for positive (POS) or negative (NEG) scores or for any of the individual relationship provisions (AFF, ALL, WOR, CON, COM, ANT, DIS, AID, NUR). Displayed in Figure 7.8, we can see that reassurance of worth (WOR) is the highest rated provision for Conditions C and D in weeks 3 and 6.

We found that across all 10 participants, there was an average of chance that they would return to do more sessions with *Mortimer* if the opportunity existed 4.44 out of 5. There was, however, no statistical difference between the groups.

7.6 Discussion

We will use the model of human-robot relationships defined in Section 3.1.3 when analysing how the results in Section 7.5 can provide answers to **RQ2** and

Table 7.3: Results of Random Intercept Linear Mixed Effect Model For Automatic Video Analysis From Groups C and D in Study 2

Condition	**Fixed Effect**	**Estimate** β	**CI [5% 95%]**	p
Robot, playing	Week	-1.37	[-2.96 0.3]	0.1764
	Group	28.75	[12.72 45.56]	0.036*
	Week.Group	4.11	[1.05 7.53]	0.015*
Robot, not playing	Week	-0.75	[-2.99 1.41]	0.5882
	Group	26.28	[14.91 38.71]	0.0125*
	Week.Group	-1.234	[-5.83 3.61]	0.0520
Piano, playing	Week	0.1283	[-0.87 1.14]	0.8436
	Group	0.68	[-6.4 7.69]	0.8906
	Week.Group	-1.028	[-3.04 1.06]	0.8736
Piano, not playing	Week	1.259	[0.44 2.15]	0.0235*
	Group	-2.933	[-7.35 1.51]	0.3638
	Week.Group	0.744	[-0.94 2.49]	0.1045
Elsewhere, playing	Week	-0.1974	[-1.35 0.95]	0.7786
	Group	-12.31	[-22 -2.81]	0.099
	Week.Group	0.0794	[-2.22 2.45]	0.3228
Elsewhere, not playing	Week	-0.973	[-2.42 0.50]	0.2819
	Group	-10.19	[-17.97 -2.39]	0.0980
	Week.Group	3.543	[0.71 6.35]	0.0535

Random Intercept Linear Mixed Effect Model for participant focus during session (%). P values are estimated from a parametric bootstrap (2000 replicates). Confidence Intervals are estimated from a parametric bootstrap (2000 replicates). *$p<0.05$, **$p<0.01$, ***$p<0.001$.

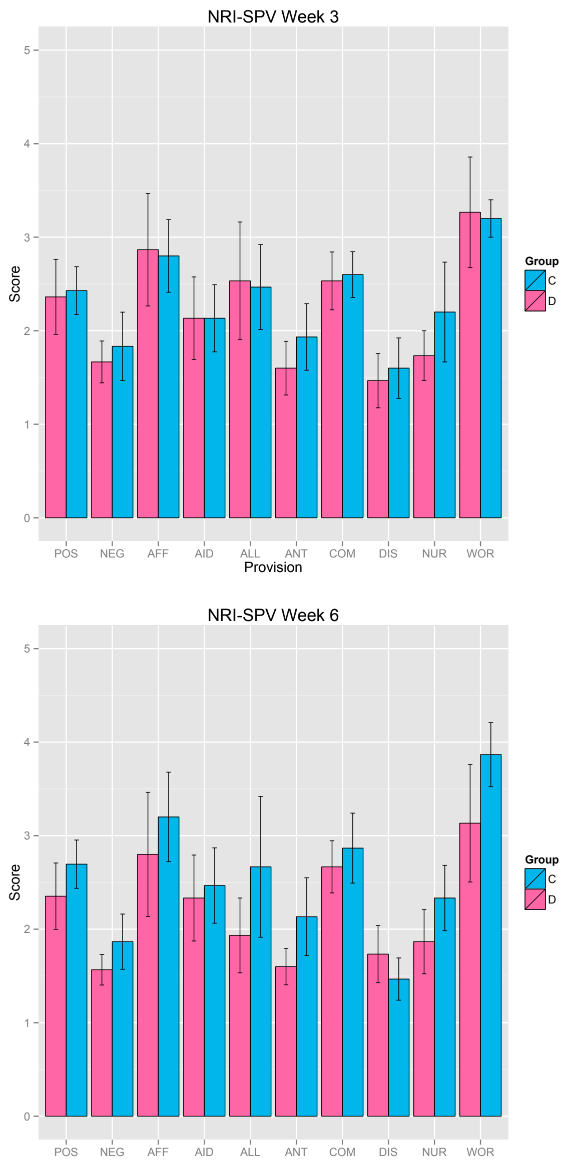

Figure 7.8: Results for C and D at Midpoint and Completion for NRI-SPV in Study 3

RQ3. As such, we will be looking for development and maintenance of social presence and engagement over multiple interactions where the human party displays some behaviours indicative of a positive interpersonal relationship.

Investigating changes in a range of quantitative interaction measures and participant focus of attention between experimental conditions and over time, we found several results of interest. Going someway to address **RQ3**, some show an effect of the control condition and demonstrate the difference that introducing head movements and facial expressions can make. Others displayed a change as the the study progressed, allowing us to draw more general conclusions in relation to **RQ2** regarding the use of music as a platform for developing human-robot relationships.

Primarily, and most crucially, we found that those in Condition C spent more time voluntarily playing with the robot over the course of the study. Demonstrated clearly in Figure 7.4, they also actually increased the time they spent as the study continued. Given Bickmore's definition of engagement as the degree of and regularity users choose to interact with the robot [Bickmore et al., 2010], we confidently take this as a sign of the positive effect of including the nonverbal behaviour. Further, its inclusion has not only avoided the novelty effect but reversed it. Users seemingly became more engaged with the robot over time.

Beyond this, we examined the way the participants used the system in both musical and social contexts during the sessions. With regards to the latter, we found that participants used the tablet to interrupt the robot less in Condition C overall, implying a greater social presence with a robot utilising nonverbal behaviour. Participants were less willing to cutoff the talking and move on as they would if using a computer program or instrument. Moreover, for both groups this decreased over time, suggesting that social presence grew as the trial progressed. These positive results were not mirrored for the button stop, the musical equivalent of an interruption, where we found no significant differences.

As musicians often use head movements as as cues during performance, especially during improvisation, we predicted nonverbal behaviour would aid the fluency of the music played, reducing frustration and aiding long engagement. However, we found longer tracks within the session for the group as a whole as time progressed, showing more engaged, uninterrupted playing. This suggests learning over time was a more important factor than the inclusion of nonverbal behaviour. Further, the finding that, regardless of group, participants spent less time playing and more time interacting socially as time passed shows that although music is the main focus of the sessions, users increasingly explored *Mortimer's* social faculties as well.

Gaze can have a large effect on the dynamics of dyadic social interaction. Mutual gaze is thought to be revealing about the interpersonal relationship between participants, for example, as a display of immediacy [Abele, 1986]. This would suggest reduced social presence in Condition C and run counter to results from the quantitative interactional data. However, Gratier does claim that mutual gaze serves less of a purpose for grounding musical interactions than it does in conversation [Gratier, 2008] so it may only be the findings of reduced focus towards the robot whilst not playing that cause concern. This being said, there is also evidence to suggest that mutual gaze occurs less as a relationship develops in social situations [Schulman, 2013], so it may be that the reduced focus is in fact a signifier of a closer relationship. Leite et al. find that children reduced the amount of time they spent looking at chess-playing robot iCat across their studies and suggest that this could be a sign of reduced social presence [Leite et al., 2009]. However, they offer the alternative interpretation that reduced attention meant that children were spending less time trying to decode the robot's behaviour. In the context of our results this could mean that the nonverbal behaviour allowed *Mortimer's* intentions to be relayed quickly, whilst those in Condition D spent more time looking to try and work these out.

We suggest the indeterminate results from the NRI-SPV demonstrate that in our case surveys lack the required sensitivity to examine human-robot relationships as it failed to find differences between the groups when the behavioural metrics showed clear effects. This strengthens our resolve that the use of behavioural metrics is the favourable approach for our interests. Based on results from Chapter 4, the high ratings for reassurance of worth (WOR) are in line with our expectations. However, this is not the case for instrumental aid (AID).

7.7 Conclusion

We uncovered several results which lead us to believe improvised musical interaction is a solid grounding for building long-term, sustainable and positive relationships between humans and robots. Again, this solidifies our resolve for a positive answer to **RQ2**. In relation to **RQ3**, our hypothesis that a positive human-robot relationship based in musical activity is aided by the the inclusion of appropriate head poses and facial expressions in both musical and social contexts is supported by our quantitative interaction data. However, this interpretation is somewhat less categorical in relation to participant gaze.

Chapter 8

Study 3: Using Online Presence to Extend Human-Robot Relationships

8.1 Introduction

In Chapter 7, we found that in a long-term study, it possible for music to provide the necessary engagement for a positive and sustainable social relationship between human and robot to develop. Moreover, that nonverbal behaviours improve the potential for a positive and sustainable social relationship between human and robot based in music. This provides compelling evidence for both **RQ2** and **RQ3**.

In our next study we examine the effect of extending the relationship beyond lab based sessions with a physical robot. One of the limitations of our previous research had been that we only have one robot which is usable in fairly supervised and regulated studio situations. Whilst it is unlikely that two humans would have complete and unadulterated access to eachothers time and location, the constraints of the current physical embodiment are more restrictive than those of a traditional relationship. We believe this to be to the detriment of engendering and maintaining long-term engagement.

Social media is something already regularly used by musicians. Nancy Baym has studied the use of social media between musicians and their fans [Baym, 2012] and between those that become friends via music based social media [Baym and Ledbetter, 2009], however, there is little research into how fellow musicians use social media between each other. She finds that although it often seen as a modern necessity of being a popular musician to engage with fans through social media, these interactions often end in real life encounters and friendships. Whilst Baym concentrates mainly on established artists, Sargent finds that local bands also use social media to find fans and develop social capital [Sargent, 2009].

Placing a robot within a human's social network can have advantages in developing a relationship. Liu et al. combine the virtual world of a 3D avatar with what they term as the "internet world" of social media sites with the aim of improving human-agent social relations. They found that a virtual human that made use of Facebook profile information to build a model of the participant's emotional state built up a better rapport and was deemed more trustworthy when offering film recommendations. The Facebots project set out with the aim of using data from an online social network to inform social interaction in the physical world [Mavridis, 2010]. One of the first to situate a robot within a virtual social network, it also used pictures from the sites to inform facial recognition [Mavridis et al., 2011]. Virtual "bots" have even been used on mobile dating app *Tinder* [BBC, 2012].

Facebook currently has 1.23 billion active monthly users [Newsroom, 2015]. Since becoming the dominant social network after the decline of forerunners Friends Reunited and MySpace, it is used by those across generations and the globe to extend their existing social relationships from the physical world into the virtual world. Although now offering a wide range of services, its original purpose as a place to share photographs online remains core to most user's experiences.

Taking the above into account, we extended *Mortimer's* capabilities to allow him to take pictures during sessions and post them with a supporting comment to Facebook. We then conducted a long-term study similar to the one described in Chapter 7 into any advantages this might have in supplementing sessions focussed around musical improvisation. This consolidated previous work concerning **RQ2** and further investigated **RQ3** by increasing our understanding of extending musical interaction with simulated social behaviours.

8.2 Design Considerations: Learning from Migration

The process of transferring a consistent agent across multiple virtual or physical embodiments is known as migration [Gomes et al., 2011]. Whilst we are not technically migrating our robot across embodiments, similar issues are raised as we are attempting to provide both virtual and physical presences to *Mortimer*. One of the first experiments into migration between virtual embodiments was the *Talking Heads* project by McIntyre et al. at Sony CSL [McIntyre et al., 1999]. Koay et al. investigate migrating a personality between two different physical embodiments [Koay et al., 2009]. In a video based study, they demonstrated that children can understand the transferring of a personality from one robot to another through a range of visual cues and suggest this implies that personality is a stronger identifying factor than its visual embodiment. However, making sure the migration is clear and smooth is important, as phenomena such as "overlaps" or "gaps" that can happen as a result of technical difficulties can cause uneasiness or anxiety [Segura et al., 2012]. Robert et al. place similar weight on the importance of consistency between virtual and physical worlds in a mixed reality robot game [Robert et al., 2011]. This may not be an issue for us as we will not be migrating between two embodiments in realtime although it is worth noting the importance of making the link between both representations of the robot clear.

8.3 Method

8.3.1 Participants

Participants were recruited by emailing musical lists and placing adverts on musician recruitment websites. There were 11 participants, 7 male and 4 female between the ages of 18 and 44. There was a wide range of self reported skill level (1-5=beginner-expert, min=1, max=5, mean=3.4, SD=1.07). Even though the number of participants is relatively small, a practical constraint of needing skilled participants, since each returned multiple times we conducted 66 sessions in total.

Table 8.1: Results of Social Media Usage Questionnaire in Study 3

Question	Min	Max	Mean	S.D.
a) How likely are you to use social media in a day?	2	5	4.08	1.24
b) How likely are you to comment a photo in a day?	1	4	1.7	0.95
c) How likely to are you post a photo in a day?	1	3	1.5	0.71
d) How many minutes per day?	30	270	91.5	78.88

1-not at all, 2-slightly, 3-moderately, 4-very, 5-extremely

8.3.2 Experimental Setup

Prior to the start of the study, participants were asked to complete a short background questionnaire with regards to their demographic information and social media usage. Results are shown in Table 8.1. From this a social media usage score was generated for each participant. Participants were then split into 2 groups, ensuring a diversity of gender and social media usage between experimental conditions. The social media usage score was calculated as $(a+b+c)*d$.

Participants were either placed in the social media interaction group, Condition E, or in the control group, Condition F. During the sessions, interactions between groups were socially and musically identical, with *Mortimer* using the same nonverbal behaviours as Condition C in Study 2.

For those in Condition E, a simple interaction based around picture sharing on the social networking site Facebook.com was enacted beyond the sessions. 3 days prior to their first session, a friend request was sent to the participant from a Facebook account purporting to be owned by *Mortimer*. The friend request had to be sent manually by the researcher as the Facebook GraphAPI does not allow this to be done programmatically. If the friend request had not been accepted by the time of the first session, *Mortimer* was programmed to politely remind the participant at the end of the session. However, all 6 friend requests were accepted prior to the first session and this was never necessary.

The information provided to participants is given in Appendix A.3.

During each session, a webcam was placed in the lab. It was located so that the shot would include both the robot and the human pianist, with 3 different

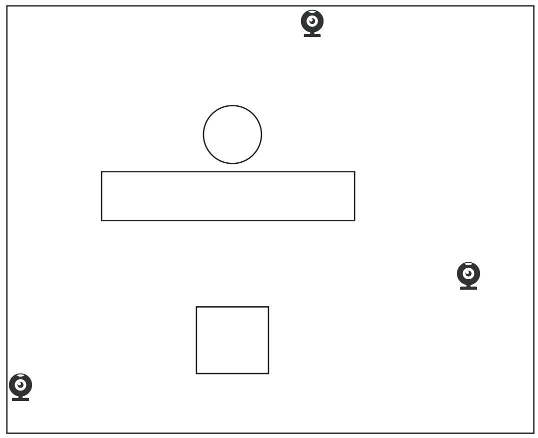

Figure 8.1: Camera Positions for Taking Facebook Pictures in Study 3

positions used throughout the first 5 sessions, varying with the week. These are shown in Figure 8.1. The picture was always taken whilst the two were playing. Once taken, the photograph was stored and a comment generated. Kasap et al. report that reference to past interactions prevented the usual decline in engagement over time [Kasap and Magnenat-Thalmann, 2011] and so this comment included either what was happening in the session at the time the photo was taken with regards to performance parameters, for example, how fast they were going, or how the session ranked comparatively to other sessions with regards to performance parameters and session length, for example, if this was the longest session to date.

Once a day a script was manually run which found any pictures that needed to be posted and added them to *Mortimer's* Facebook wall. It also included the comment, with an optional further comment about weather at the time of the session [1]. The user was also tagged in the photo, meaning they received a notification of the photo's posting. The script also checked for any interactions from users since it had last been run, such as likes, comments or messages, and stored this in order to thank the user for their contact in any proceeding comments.

The script also checked for any comments or replies to the photos it would post. They would then be run through the sentiment analysis algorithm of Narayanan, Arora and Bhatia [Narayanan et al., 2013] and if they were deemed positive or neutral, *Mortimer* would post a reply thanking them. This would

[1] http://openweathermap.org/api

Figure 8.2: An Example of an Automatically Generated Post to Facebook by Mortimer during Study 3

allow a convincing reply but not require any natural language processing as the to the meaning of the comment beyond ascertaining whether the phrase was broadly positive or negative, reducing the chance for incorrect interpretation. Once an interaction on a particular thread had been replied to once, *Mortimer* would ignore any further interactions, such as replies to his reply, to avoid beginning a conversation. A simple reply thanking the user serves a similar purpose to the "like" facility provided by Facebook, however, the GraphAPI does not allow things to be liked programmatically. The GraphAPI also does not allow messages to be sent programmatically, so *Mortimer* was unable to reply to any private messages sent to him.

8.3.3 Measures

Taking the methodological approach detailed in Chapter 3, we measured both quantitative interaction data relating to the way the participant's interacted with the robot socially and musically and used Soyel and McOwan's facetracking algorithm to determine their focus throughout the sessions. The measures taken are the length of the session over the minimum 20 minutes, the mean number of bars per track, the percentage of the session spent playing the piano, the number of explicit "button" stops and the mean number of interruptions per session.

We also administered the NRI-SPV questionnaire, modified for use with robots, at the mid point and after the final session. As a measure of intended repeat interaction, as part of the exit questionnaire participants were asked how likely they would be to attend if they had the chance to have more sessions on a Likert scale (1-5).

Additionally, we recorded the number of "likes" and comments by participants and members of their social network on posts made by *Mortimer* and posts made by themselves that they tagged *Mortimer* in. "Likes" are useful in our quantitative approach as they are by definition a positive interaction, whereas this is not necessarily the case with a high number of comments. We were unable to record any references to the sessions that did not explicitly involve *Mortimer's* facebook account for privacy reasons. Also, as the use of Facebook is the experimental condition, only Condition F could provide data and so no comparisons could be made between the groups.

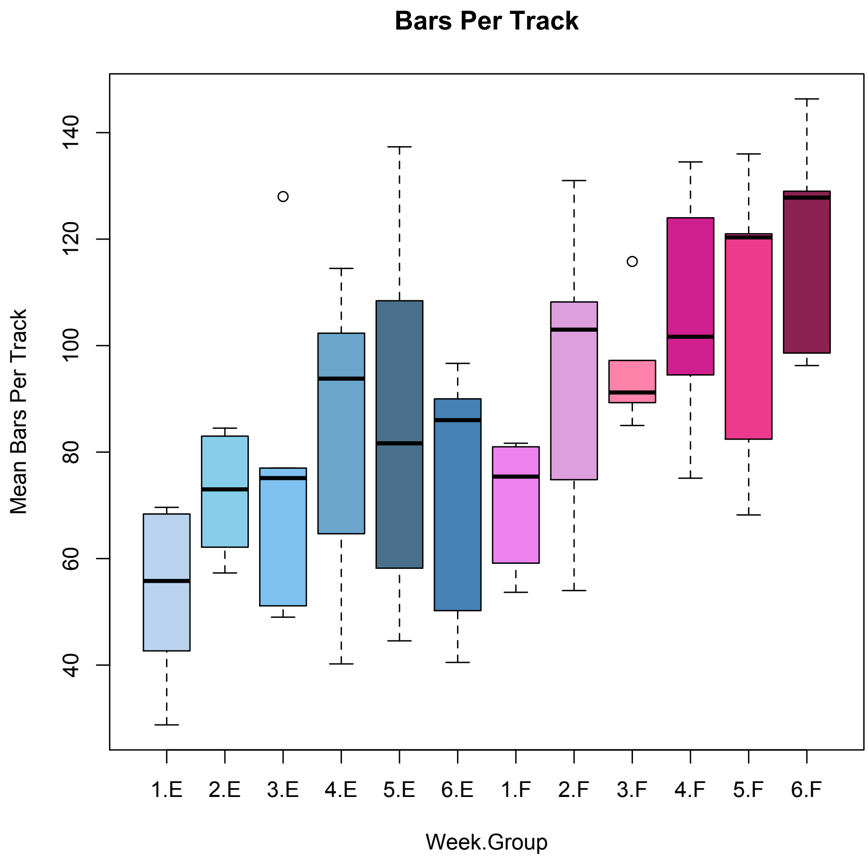

Figure 8.3: Mean Bars Per Session for Groups E and F in Study 3

8.4 Results

8.4.1 In Session Results

Quantitative Interaction Data

As in Chapter 7, we fitted a random intercept linear mixed effect model for the fixed effects of week, group and the interaction between the two for each measure. Results are displayed in Table 8.2.

We found significant effect of week ($\beta = 6.21$, 95% CI [4.19 8.27], p=0.0005) and the interaction between week and group ($\beta = 3.89$, 95% CI [0.01 7.79], p=0.0005) for the mean number of bars per track. Shown clearly in Figure 8.3,

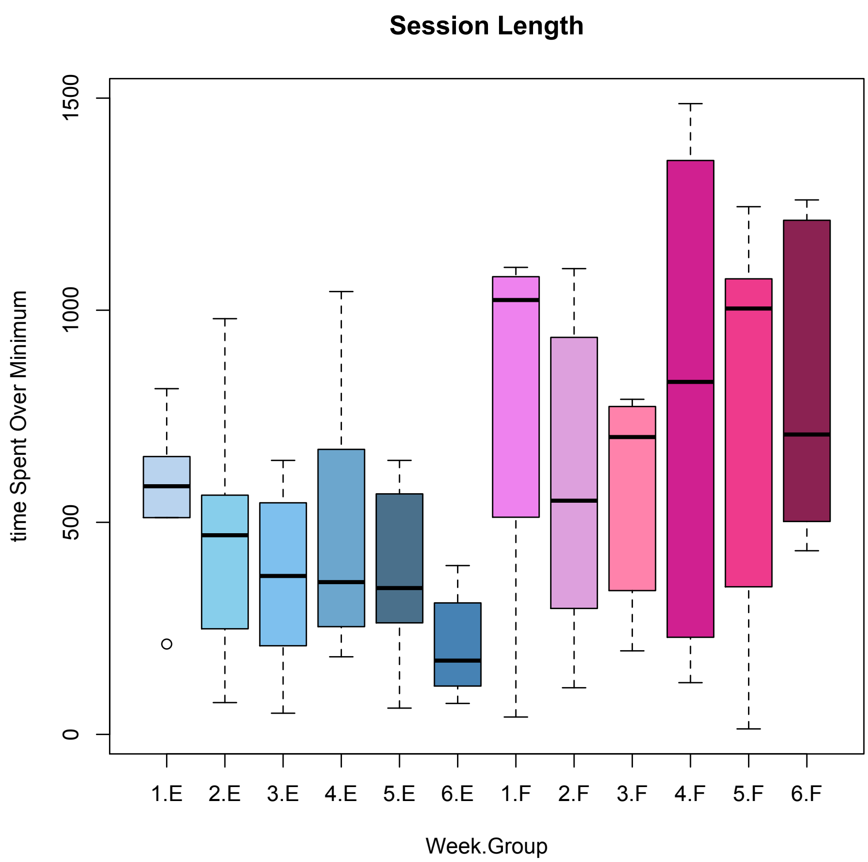

Figure 8.4: Session Length for Groups E and F in Study 3

Data	**Fixed Effect**	**Estimate β**	**CI [5% 95%]**	p
Session	Week	-17.11	[-44.57 10.68]	0.3193
Length	Group	303.63	[36.59 591.65]	0.1229
	Week.Group	84.60	[33.13 135.78]	0.0150*
Bars Per	Week	6.21	[4.19 8.27]	0.0005***
Track	Group	24.10	[7.97 40.53]	0.0605
	Week.Group	3.89	[0.01 7.79]	0.0005***
Button	Week	-0.01	[-0.02 0.00]	0.2584
Stops	Group	-0.07	[-0.2 0.08]	0.4968
	Week.Group	-0.02	[-0.05 0.00]	0.3398
Inter	Week	-0.04	[-0.11 0.04]	0.4343
-ruptions	Group	-0.01	[-0.64 0.54]	0.9750
	Week.Group	0.12	[-0.03 0.27]	0.5567
Time	Week	0.71	[0.01 1.40]	0.0960
Playing	Group	4.60	[0.83 8.39]	0.1034
(%)	Week.Group	-0.40	[-1.79 0.93]	0.1109

Random Intercept Linear Mixed Effect Model for quantitative interactional data. P values are estimated from a parametric bootstrap (2000 replicates). Confidence Intervals are estimated from a parametric bootstrap (2000 replicates). *p<0.05, **p<0.01, ***p<0.001.

Table 8.2: Results of Random Intercept Linear Mixed Effect Model For Quantitative Interaction Data From Groups E and F in Study 3

the mean number of bars per track increases as the weeks continue, and whilst there is no significant difference between the groups overall, the change over time is greater for Condition F.

We further find an effect of the interaction between week and group (β = 84.60, 95% CI [33.13 135.78], p=0.0150) for the time spent of the minimum in each session, indicating that Condition F has a positive effect on the trend over time in relation to Condition E. In this case, Figure 8.4 demonstrates the steady reduction in time, over time, for those in Condition E, whilst those in Condition F see a week on week rise for week 2 to 5 before a drop off in the final session.

Condition	Fixed Effect	Estimate β	CI [5% 95%]	p
Robot, playing	Week	-0.93	[-2.62 0.79]	0.3818
	Group	-3.42	[-20.26 13.23]	0.7816
	Week.Group	1.53	[-2.05 4.91]	0.7396
Robot, not playing	Week	0.39	[-1.46 2.33]	0.7426
	Group	4.82	[-38.30 -14.16]	0.6917
	Week.Group	-2.03	[-5.92 1.83]	0.8186
Piano, playing	Week	-0.26	[-1.21 0.63]	0.6452
	Group	-0.85	[-10.62 8.92]	0.9135
	Week.Group	0.78	[-0.99 2.7]	0.8931
Piano, not playing	Week	0.54	[-0.47 1.58]	0.3688
	Group	-6.01	[-16.85 4.14]	0.4098
	Week.Group	0.6	[-1.38 2.53]	0.6277
Elsewhere, playing	Week	1.55	[0.32 2.84]	0.0555
	Group	0.31	[-4.52 4.91]	0.9155
	Week.Group	-0.19	[-2.76 2.48]	0.2909
Elsewhere, not playing	Week	0	[-1.4 1.41]	0.9975
	Group	0.33	[-4.55 5.19]	0.9215
	Week.Group	0.7	[-2.16 3.48]	0.9785

Random Intercept Linear Mixed Effect Model for participant focus during session (%). P values are estimated from a parametric bootstrap (2000 replicates). Confidence Intervals are estimated from a parametric bootstrap (2000 replicates). *$p<0.05$, **$p<0.01$, ***$p<0.001$.

Table 8.3: Results of Random Intercept Linear Mixed Effect Model For Automatic Video Analysis From Groups E and F in Study 3

Automatic Video Analysis

As in Chapter 7, we fitted a random intercept linear mixed effect model for the fixed effects of week, group and the interaction between the two for each category. Due to a technical fault, video was unavailable for analysis for one session from the first week. Therefore, the analysis is done on 65 of the 66 sessions. The results, displayed in Table 8.3, show that there was no significant differences between the groups or over time for any of the categories.

Self Report

We conducted T-tests for each provision and the amalgamated positive and negative scores. Analysis of results did not find any significant factors between groups or over time for positive (POS) or negative (NEG) scores or for any of the individual relationship provisions (AFF,ALL,WOR,CON,COM,ANT,DIS,AID,NUR).

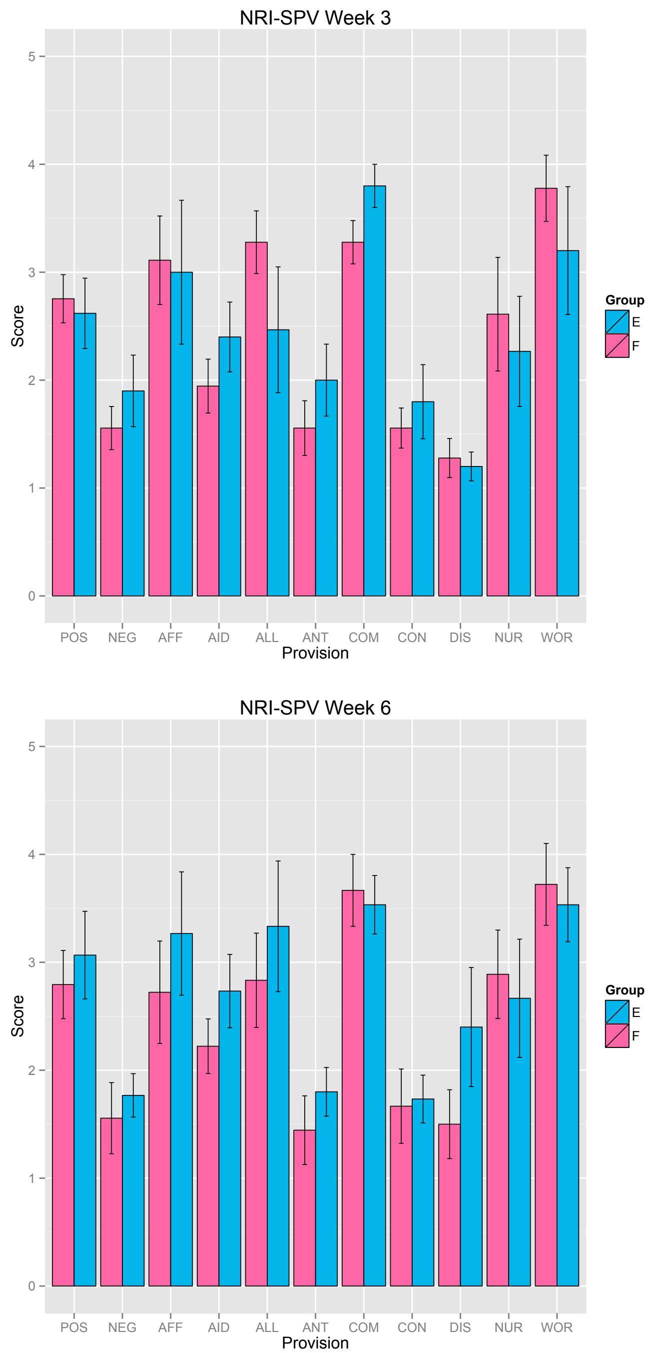

Figure 8.5: Results for E and F and Midpoint and Completion for NRI-SPV in Study 3

Displayed in Figure 8.5, our results show similar findings to Chapter 7. As predicted by our results in Chapter 4, for both groups and weeks, reassurance of worth (WOR) was the amongst the highest rated provision, along with companionship (COM).

We found that of all 11 participants, they reported an average of 4.45 out of 5 chance that they would return to do more sessions with *Mortimer* if the opportunity existed. There was, however, no statistical difference between the groups.

8.4.2 Facebook Interaction

As by far the most commonly occurring form of interaction with posts, we have displayed the likes by others in Table 8.4. It shows that the total number of likes a post received was considerably higher if the user posted the picture themselves, although this only happened twice, both times in the first week. It also demonstrates that posts by *Mortimer* did not suffer a novelty effect in the participant's social networks, as the highest proportion of posts liked (5) and the second highest mean number of likes per post (2.17) occurred in the final week.

Likes by participants themselves were rare, with only 3 occurring over the whole 5 weeks. Comments by participants only occurred once once comments by others only 3 times. In contrast, the two posts by participants received 1 and 6 comments, making comparative means of 0.1 and 3.5 respectively.

8.5 Discussion

In relation to **RQ3**, based on previous studies and the overarching hypothesis running throughout this thesis, we would expect the inclusion of additional social modalities to result in increased engagement. We have suggested that the length of time a person voluntarily spends with a robot can be a key indicator of engagement and in Chapter 7 reported that those in the social condition not only spent more time overall but also increased the amount of time they spent as the study continued. A repeated trend is seen in this experiment, with the time spent increasing over the study (see Figure 8.4, Conditions C and F are analogous). Further, we find a significant difference in how the session length

Table 8.4: Number of Likes Received From Non-participant Users for Facebook Posts From Put Online by Mortimer and Participants During Study 3

Week	**Group**	**Posts Liked**	**Min**	**Max**	**Mean**	**S.D.**
1	Par	2 (of 2)	22	63	42.5	28.9
1	Rob	4 (of 6)	1	3	1.5	1.38
2	Par	0 (of 0)	0	0	0	0
2	Rob	4 (of 6)	1	5	2.33	2.25
3	Par	0 (of 0)	0	0	0	0
3	Rob	1 (of 6)	2	2	0.34	0.82
4	Par	0 (of 0)	0	0	0	0
4	Rob	2 (of 6)	1	6	1.17	2.40
5	Par	0 (of 0)	0	0	0	0
5	Rob	5 (of 6)	1	6	2.17	2.04

Par=Participant Posted, Rob=Robot Posted

changes over the study between the two conditions, however, it is the more social condition, Condition E, that actually reduces over time.

With a study design that allowed identical access to the robot for all conditions, as with Study 1 and Study 2, this result would appear to show a decrease in engagement for those in Condition E, however, this was not the case. Participants in Condition E had opportunities for additional contact with *Mortimer* outside of the physical sessions, possibly leading to a reduced need to spend time in the physical sessions. This result highlights the potential issues that may arise from study design when picking experimental measures and illustrates the difficulty in providing an answer to **RQ1** that works across all possible HRI contexts. In this case, a measure suitable in previous studies is compromised by new modes of interaction. It also raises an interesting question regarding whether humans will spend less time with a physically embodied robot if they know they can interact with it later on a virtual platform.

Similarly to Study 2, the length of tracks within sessions increased over the trial period, regardless of experimental condition. We can again draw the conclusion that learning over time is a more important factor than increased social modality in increasing the fluidity of playing.

From the Facebook data, there were considerably more ”likes” from a user’s social network for posts made by a user themselves, as opposed to one posted by *Mortimer* that the user was tagged in. There are several explanations for

Figure 8.6: An Example of a Facebook Post Including Mortimer By a Participant

this outcome, one being the differences in the photos themselves. As seen in Figure 8.6, both participant posted pictures were "selfies" in which the participant is facing the camera and engaging the audience. This is in contrast to the photographs posted by *Mortimer*, taken whilst playing and often without the player looking directly at the camera (see Figure 8.2). Alternatively, this could have been caused by the quality of the comments generated. It could be that *Mortimer's* comments were viewed as prescriptive or disingenuous and the participant's were more natural and "likeable". It may even have been a bias against the poster that can explain the disproportionate amounts of "likes". For example, just the fact the picture was posted directly by a friend as opposed to a stranger, in this case *Mortimer*, or that one was posted by a robot may have affected people's decision to like the post to or not. One thing to note is that proper comparisons between the two groups, self posted and robot posted, are limited due to the data set having 30 entries in the latter and only 2 in the former. This makes any generalisations hard, although the size of the difference, even with the limited data, makes it worthwhile considering.

Unlike in Chapter 6 and Chapter 7, where facetracking revealed differences in the focus of the participants, no differences were found between the groups in this experiment. However, unlike the previous studies, the in-session interaction between the groups was identical in this case. From this it may be seen that any changes caused by the experimental condition are less immediately reflected in the nonverbal behaviour of the participant.

Again, there was a generally high score for self-reported repeat interaction, suggesting that the although the experimental condition may not have had the expected influence, the insession activities were engaging for all participants. This whilst providing more evidence for an affirmative answer to **RQ2**, it takes considerably more effort to actually stay and interact than to tick a box suggesting you would, so actual voluntary session times are a taken as a much stronger signifier than self report in this case.

8.6 Conclusion

Using the methodology of automated behavioural metrics developed in Chapter 3, we found the effects of extending the relationship into the virtual world were less pronounced than results we have previously found by adding social modalities to human-robot musical interaction. With regards to **RQ3**, this suggests that some simulated social behaviours will have a greater effect than others on improving the potential for a positive and sustainable social relationship between human and robot based in music. With implications for **RQ1**, the results also raised a question as to the appropriate use of session length as a measure of engagement in this context. Further, analysis of the Facebook data provided some noteworthy differences in interactions with posts by participants and posts by *Mortimer*, however, the former category had too small a dataset to draw any solid conclusions. Continued long sessions throughout and positive reports of self intended repeaet interaction again provide strong support for a positive answer to **RQ2**.

Moving forward, more experiments would illuminate whether extending the relationship into the virtual world is simply not a particularly useful tool in this context or that a higher quality of interaction is required to trigger positive effects on human-robot relationships.

Chapter 9

Conclusions

To conclude this thesis we will first recall the thesis statement presented in Section 1.1.

Thesis Statement In this thesis, we will examine whether the addition of simulated social behaviours will improve a sense of believability or social presence, which, along with an engaging musical interaction, will allow us to move towards something that could be called a human-robot relationship.

After surveying the existing research fields relevant to this statement in Chapter 2, we highlighted underexamined gaps and challenges in these areas. Taking these into account, in Section 2.4 the thesis statement was neatly unpacked into three individual research questions which this work sought to address. Below we will return to each research question individually and consider to what extent it has been answered by the novel work contained in this thesis. In doing so we will also shed light on the the original question posed in the thesis statement.

- **RQ1** How can we determine the quality of a social relationship developed between and human and robot over multiple social interactions?

This was addressed in Chapter 3. After identifying the necessary factors that make up a human-robot relationship, the following model was proposed:

A human-robot relationship can be defined as the development and maintenance of social presence and engagement over multiple interactions where the *human party displays some behaviours indicative of a positive interpersonal relationship.*

Eschewing the prevalent approaches of self-report, we presented a methodology of automated behavioural metrics to evaluate the quality a human-robot relationship based on the model above. Although not formally validated, this model and methodology are resuable for other HRI researchers looking to design and evaluate robots built for long term social interactions. Proceedingly, we asked:

- **RQ2** Is it possible for music to provide the necessary engagement for a positive and sustainable social relationship between human and robot to develop?

Primarily, the literature review in Section 2.2.1 provided strong evidence in support of **RQ2**. This was further supported by results from a large online study in Chapter 4. Using the well-validated NRI-SPV survey, we found that there were statistical equivalences for a key group of relationship provisions that participants perceived to be engendered by both friends and people they played music with regularly. Namely, these were reassurance of worth (WOR) and instrumental aid (AID) or guidance.

In all three HRI experimental studies carried out with *Mortimer*, the independent variables have been the inclusion or not of incrementally advanced simulated social behaviours. As such, the effect of music cannot be analysed with the statistical comparison of experimental and control groups. Nonetheless, music was the main activity of all the sessions. In Study 1, 90% of participants stayed for the full session length, despite being able to leave at any point. Further, in Studies 2 and 3, some participants were still staying for over twice the minimum time 6 weeks into the study. The exit questionnaires of these studies also showed the mean score for self reported chance of continuing with more sessions if possible was 4.45 out of 5 (89%). This shows that, regardless of the simulated social behaviours, musical activity was able to provide both an initial engagement and maintain this engagement over time. This clearly supports an affirmative answer to **RQ2**. Finally, a third question was posed:

- **RQ3** Which social behaviours improve the potential for a positive and sustainable social relationship between human and robot based in music?

The results from the online survey in Chapter 4 provided us with high level provisions that participants would expect to receive from a relationship built around joint musical activity. These were then developed in Section 4.5 as considerations to be used when designing the social behaviour of *Mortimer* to best best engender a positive human-robot relationship. For example, to be good reassurers of worth, *Mortimer* flattered participants when they succeeded and refrained from not admonishing them when they failed.

RQ3 is clearly addressed by Studies 1, 2 and 3 in which social behaviours are incrementally evaluated in controlled experiments where humans and robot improvise music together. Although a single session experiment, Study 1 provided clear evidence that framing a musical interaction between human and robot as a social interaction increased both engagement and social presence, two major factors in our proposed model of a human-robot relationship. Study 2 again shows increased social presence and engagement with the inclusion of socially and musically triggered head poses and facial expressions. As a long term trial, we were also able to uncover a compelling reversal of the novelty effect. In Study 3 we found the effects of extending the relationship into the virtual world were less pronounced than results we had found previously. With regards to **RQ3**, this suggests that some simulated social behaviours will have a greater effect than others on improving the potential for a positive and sustainable social relationship between human and robot based in music.

The sections below cover the contributions to the body of academic knowledge by this research. Moreover, possible avenues of research opened up by this work are detailed.

9.1 Research Contributions

Although a robot drummer has been built and composition algorithms developed, there is no claim to either of these being state-of-the-art in their respective fields. The most important contributions of this thesis are those that directly address the research questions posed in Section 2.4. Namely, a model and methodology for examining human-robot relationships and the findings of long-term HRI studies using this approach.

This research has been presented at well respected, peer reviewed conferences to engage directly with the HRI and Social Robotics communities. It has also been involved in a wide-range of public engagement activities, dispersing the interest in the potential of Social Robotics through science festivals, national television and live comedy shows to audiences well beyond the normal academic spheres. That these results did not depend on state-of-the-art technology itself highlights that the solution to having and maintaining positive and engaging social interactions with robots is not necessarily in the most sophisticated AI and engineering techniques but in well designed and well researched behaviours and activities. For example, our use of open ended creative tasks. Further, although the technology in itself was limited, to us it was important that experiments were conducted with actual functioning technology rather than taking a WoZ approach.

9.1.1 A Methodology for Evaluating Human-Robot Relationships

In **RQ1** we queried how the quality of a social relationship developed between and human and robot over multiple social interactions could be determined. The model and methodology proposed in Chapter 3 clearly addressed this. Having identified long-term engagement and social presence as key factors in maintaining a human-robot relationship, we have been critical of the questionnaire based approach taken by the majority of HRI researchers. Primarily, we feel in most cases that humans cannot accurately report the occurence of these psychological phenomena on a 5 point Likert scale. They are further inappropriate for use in human-robot relationship trials because of known biases against self reporting social behaviour with robots [Reeves and Nass, 1996] and as repeated measures may be susceptible to a learning or boredom effect. In contention to this we have developed an alternative approach of automated behavioural metrics. This purports that by identifying and measuring the occurence of phenomena that are known signifiers of either engagement, social presence or a positive interpersonal bond, insight can be reliably gained into the quality and development over time of a human-robot relationship.

The limitations of this approach lie within the task specific identification of phenomena, however, the model of a human-robot relationship and the methodology in general are still largely adaptable for use by other researchers in the HRI field. For example, it could easily be tailored for long-term trials of service, healthcare or educational robots. Adoption of this methodology within the community will help to standardise findings and make the benchmarks for progression clearer. Further, a commitment to studying the long-term effects of

Social Robotics is crucial if legitimate progress is to be made towards any type of robotic social companion and the availability of a model of and methodology for human-robot relationships will hopefully encourage further research in this area.

9.1.2 Evaluations of Engagement and Social Presence in Human-Robot Musical Interaction

Over the course of this research project, we have conducted 136 sessions with *Mortimer*, as well as a sizeable online survey. The research carried out and the results uncovered in Chapters 4 and 6 went largely to confirming our informed predictions that music can be a solid grounding for a social relationship and that the inclusion of social behaviours ontop of musical ability would be advantageous. This guided our design of *Mortimer* and emboldened us to execute the long-term studies described in Chapters 7 and 8. In doing so we addressed not only the latter two of our own research questions but also a large gap in the existing HRI literature. Below we detail our key findings.

- Statistical equivalence (δ=5%) for the provisions of reassurance of worth (WOR) and instrumental aid (AID) between human friends and regular co-musicians.
- More natural stops and longer uninterrupted playing when human-robot musical improvisation is framed as a social interaction.
- Longer sessions overall and a positive trend in session length over time, less interruptions, greater increase in uninterrupted playing over time and less looking at the robot when nonverbal behaviour is introduced in a long-term human-robot musical improvisation study.
- Negative trend in session length and greater increase in uninterrupted playing over time when virtual presence is introduced in a long-term human-robot musical improvisation study.
- Out of 21 participants completing 6 sessions with *Mortimer*, the mean score for self reported chance of continuing with more sessions if possible was 4.45 out of 5 (89%).

Within the Social Robotics community researchers rarely commit to long-term studies and as such, it is hard to extrapolate any positive outcomes beyond single encounter HRI contexts. Whilst appropriate for some robots, such findings cannot inform us about how a relationship may develop in a social context where multiple encounters are likely, such as domestic or care robots. Not only have we bridged this gap by conducting two long-term studies, but our choice of musical improvisation as a domain has uncovered results in contention to normally seen novelty effects. For example, in both studies, even by the sixth session, some participants were still staying and playing with *Mortimer* for 45 minutes, over twice the mandatory time. Significant experimental findings included an upwards trend in session length and a reduction in interruptions across 6 session studies with the inclusion of nonverbal behaviours. These findings are instructive both to the HRI community, providing insight into the use of musical improvisation in forming human-robot relationship, but also to the IMS community, who gain knowledge about the advantages of including social modalities in interactive performance and composition systems.

In reference to **RQ2** and **RQ3**, through the studies carried out above we can confirm music as a solid grounding for human-robot relationships. Also, that these initial foundations can be developed by framing the sessions as a social interaction and triggering appropriate head poses and facial expressions in both musical and social contexts.

9.1.3 Public Engagement

One of the largest hurdles in the proliferation and success of Social Robotics is its acceptance by the wider public. These are the people who will be served by them, work alongside them in their daily lives and purchase them for their homes. Any new technology, as an unknown, comes with distrust and, not helped by the avid imaginations of science fiction authors, this is especially prevalent in AI. By actively engaging the public in ongoing research we are not only beginning the slow process of exposing them to what might be but also being transparent of our motivations and openly involving them in this dialogue as we move forward.

During this research project, *Mortimer* has appeared twice on national television. Once as part of the Royal Institute's Christmas Lectures and once on technology magazine programme *The Gadget Show*. He has been met in person by many children at the Brighton Science Festival and many adults at the Royal Institute's Lates event and presented by the author at the British Science Festival as part of Professor McOwan's Presidential Lecture. *Mortimer* even made a stage debut as part of Festival of the Spoken Nerd's comedy show where issues

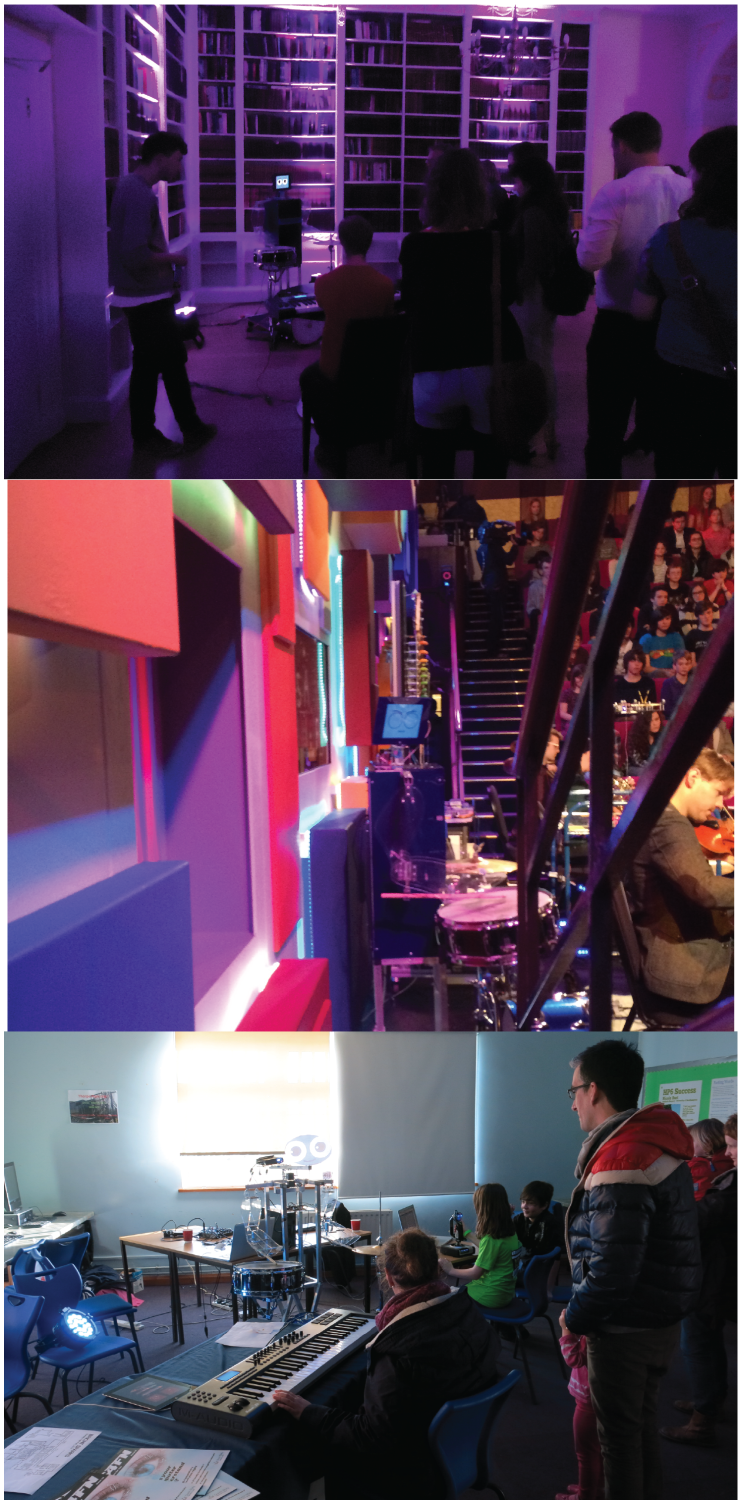

Figure 9.1: Photographs of Mortimer Engaging in Public Engagement Activities

surrounding the research were discussed in a more playful light. All events allowed people to directly interact with cutting edge science and to ask questions of and be answered by the researchers involved.

Moreover, talking to the public during these events provided invaluable insight for the author into how people used *Mortimer* and what they expected him to be able to do. This fed back into the ongoing design and development process. For example, in terms of the musical interaction, the extra question of "Are you ready?" and the one bar count to the beginning of tracks was introduced after we noticed the confusion often caused when people were not prepared for *Mortimer* to start playing immediately. Further, we were able to see whether nonverbal behaviours were interpreted as we intended. One issue uncovered that was a facial expression designed to demonstrate concentration was sometimes misinterpreted as a display of anger or unhappiness towards the pianist playing. This went against our aim to convey a positive, supportive personality for *Mortimer* and we were able to correct this for future experimental trials.

9.2 Limitations

In order to achieve the required depth and detail, any PhD thesis necessarily has a narrow scope. As such, we will now reflect on the limitations of the research carried out in this work in achieving its stated aims in terms of the thesis statement and research questions provided. The limitations of the scope of the studies in a wider research context will be explored proceedingly in Section 9.3.

A problem common to any scientific research requiring human participation, and so the majority of HRI trials, is recruiting the requisite number of participants to achieve the necessary statistical power when analysing results. For us, this problem is magnified twofold. Firstly, attending every session in a long-term trial requires a large commitment and so participants are often less willing to do so. Morever, our studies required skilled participants, in this case pianists. This reduces the pool of potential participants and again increases the challenge of recruitment. Although our studies had low numbers of participants, as they were multi-session trials multiple data points were recorded for each participant. This being said, more participants would have increased the statistical power of our analysis and so our confidence in any findings.

As we mentioned in Section 3.1.3, the scope, variance and individuality of social relationships is considerable and so any model that attempts to encompass all possible occurrences could be placed somewhere between ambitious and naive. As such, any conclusions drawn from our findings are unlikely to be universally applicable. Further, most phenomena in Social Psychology are inherently difficult to measure. This includes phenomena used in the model of human-robot relationships proposed in this thesis, for example, social presence. We have done our best to minimise this by avoiding the problematic approach of self report and taking automated measures of observable phenomena. However, it is worth noting that even the well informed interpretation of these phenomena as signifiers of a human-robot relationship is still an interpretation.

9.3 Future Work

Before moving on to future research directions, we take some time to reflect on what the implications of further developing robots capable of forming relationships with humans would be. Would these be seen as different to human relationships? Would these begin to change the way we see our relationships with other humans? Even if the feelings are genuine from the human, this will not be the case for the robot. It is, and feasibly will always be, a simulation [Dautenhahn, 2014].

With the developments of new digital technologies, on the whole consumers have chosen convenience and value over richness and resolution when given the choice. The music quality that comes from the speakers on a laptop is often considerably lower than that of even a low-end hi-fi, yet out of convenience this is increasingly the listening choice of many. The quality of streamed video is considerably less than DVD or Blu-Ray formats, yet most people would prefer to have the options of millions of movies online in lower resolution over a much smaller library of high quality films. Although choices of media consumption are by no means directly analogous to choices of social interaction, they are raised as examples where quality is sacrificed for convenience in the context of digital technology. A more fitting comparison comes with pornography and concerns that increased exposure can affect people's attitudes towards physical relations with other humans. As with social robots, it is simulation that can provide some of the gratification of the "real" thing but without any of the challenges of interacting with another biological entity. In the context of Social Robotics, if this trade-off is the want of humanity, there is little we can do to stop it, if it even is our place to admonish it. However, we feel it is our responsibility as HRI researchers to avoid providing an easy simulation at the cost of human social interaction and rather to attempt to develop systems that either directly address areas of need, for example, the loneliness of the elderly, compliment existing social relationships or allow for social and creative experiences beyond those that could be garnered from human-human interaction alone.

This thesis has dealt with the testing of a small set of social behaviours in a specific domain. Insights gained from this are intended to guide the construction of more holistic robotic systems for long-term social interaction with humans and so the potential additions of behaviours or functionalities are almost as endless as those within the Artificial Intelligence and Engineering fields. Technologies are detailed below that could provide immediate improvements to either *Mortimer* or other robots operating in similar domains.

9.3.1 Extension of Current System

Machine Learning

In the context of AI, machine learning refers to the use of data to adapt an agent's internal model. As such, its inclusion into either the musical or social intelligence of *Mortimer* could be used to address two key factors of long-term human interaction, namely, personalisation to individuals and adaption over time. MacDorman and Cowley make the case for a robot's personality and beliefs, though simulated, to follow that of the humans it interacts with and adapt alongside its relationships. In summary, if they are not able to have personally specific and developing relationships the ability to imitate human behaviours may not endow a robot with the humanity necessary to develop social relationships with humans [Macdorman and Cowley, 2006]. One of the reasons music was chosen was that it allowed for a large amount of natural progression for the participants even with a static system. However, if the robot was able to use data from previous sessions to adjust its composition algorithm as the relationship developed it could potentially avoid any declining engagement as humans explored and reached the limits of a nonadaptive system. Moreover, it would allow for a robot to begin with a reasonably agnostic style palette and hone itself more specifically to a partner's style over time.

By definition, machine learning requires data. As a by-product of our methodological approach, *Mortimer* already collects a large amount of quantitative data about how the pianists use the system and how they behave during the sessions.

However, we do operate from a cold start and so personalisation necessarily has a lead period as enough data is collected. Following this, changes in musical style or social behaviour can then be tracked and reacted to appropriately.

The most informative data for a composition system would be the transcripts of the music itself and the participant's use of performance parameters. These could inform adaption to style in the longer term and aid immediate interactivity by improving how *Mortimer* predicts upcoming music input. An example of this is the work carried out by Evana Cristina Dos Santos as a summer project for the Science Without Borders programme. Using feedforward neural networks and piano transcripts from Study 2, she was able to predict the binary rhythm of a bar based on the previous bar to an error of 0.11.

Although few immediate returns were reported in Study 3, by placing *Mortimer* within his partner's virtual social network, historical data about social activity could initially be made use of to avoid a cold start. It would also allow for up-to-date information about changes in their life to be used as the relationship developed.

Music Information Retrieval

Music Information Retrieval (MIR) refers to the automated abstraction of meaning from music. The most recent call for articles from the community's leading conference, ISMIR 2015, asks for papers on the extraction of melody, harmony, chords, timbre, rhythm, beat, tempo as well as automated categorisation and summarisation of pieces. The more information *Mortimer* can gather from a human's playing, the better he will be able to react accordingly and the better he will provide an increasingly engaging interactive musical experience. Further, creative acts remain exterior to the tasks people expect a computer to be proficient in. This means the greater the perceived creative autonomy of an artificial agent, the greater the believability and social presence it provides. All of the above techniques would aid *Mortimer* in these regards, although online beat tracking is the one we feel would provide the most immediate improvements.

Online beat tracking refers to the inference of tempo and beats from either a symbolic or raw audio stream in real time and presents a harder challenge than its reasonably well solved offline counterpart. The reason we include this is because one of the first expectations of *Mortimer* from musicians is that he will be able to match their tempo and keep in time as they change over time. Further, as a percussionist, temporal synchrony is a critical skill. Early developments of our current system integrated Adam Stark's BTrack system [Stark, 2011], however, bearing in mind our specification for a stable and reliable system, we encountered problems such as misclassification of tempo and occasional misinterpretations of correct tempo at double and half time at unacceptable regularity. This notwithstanding, the largest issue was that in the context of a percussion and piano duo, the pianist is expecting the drummer to keep a steady beat to which they will synchronise their own playing. If the drummer is matching their tempo and they are matching the drummer's tempo, a gradual slow down occurs as is sometimes seen with orchestras and unexperienced conductors. If the problems mentioned could be solved, online beat tracking would be greatly improve the feeling of responsiveness of the system, creating a more seamless musical experience.

9.3.2 Different Systems

It may be that the positive results from this research may apply to other forms of creative practise, yet few provide the real time collaborative advantages of musical improvisation. One of the only activities to share this is dance. Several projects have involved the use of solo dance by a robot as an expressive behaviour and as a visual display of musical synchrony [Hoffman and Vanunu, 2013, Grunberg et al., 2009, Nakahara et al., 2009]. Michalowski, Sabanovic and Kozima even explicitly use rhythmic syncing to background music as a way to promote social interaction [Michalowski et al., 2007]. Others use pressure sensors [Michalowski et al., 2009] or cameras to synchronise dance between humans and robot in real time. We think these systems could do well in a Social Robotics context as postural synchrony and close proximity are actively encouraged and these are often reported as a signs of a close interpersonal relationship. An intriguiging future research direction would be to see if similar long engagement results to ones achieved in this thesis could be seen with improvisational, interactive human-robot dance. Beyond this, another important advantage that could be gained from the use of human-robot dance would be the intimacy that comes with tactile interaction. Although the coordination with a human in real-time makes this a much harder techincal task than the projects reported above, a ballroom dancing robot is described by Takeda, Hirata and Kosuge [Takeda et al., 2007] using a HMM approach to predicting the next dance move in a waltz from a male leading dancer. Again, it would be interesting to see how this approach faired in a less prescriptive situation, over multiple sessions or with the inclusion of other social modalities.

9.4 Closing

In this thesis we investigated the effectiveness of musical improvisation as a means of providing the necessary engagement for human-robot relationships to develop. Our results provide us with confidence in this approach. Further, by experimenting with various simulated social behaviours, we have also found evidence suggesting that this approach may be suited to fit into more holistic companion robots.

We hope this work highlights the importance of investigating how social interactions between social robots and humans can progress or deteriorate over time and the role that open-ended creative activities can take in avoiding a decline in favourable response, aiding the development and maintainence of positive and sustainable human-robot relationships.

Appendices

Table of Contents

Appendix A

Forms

A.1 NRI-SPV for Robots

A.2 Ethics Approval

As all studies included human participants, the research code of Queen Mary, University of London requires formal ethics approval. This was gained for all studies, with the appropriate reference numbers listed below.

- Online Study - QMREC1030 Musicians and Companionship
- Study 1 - QMREC1299 Social Dialogue in Human-Robot Musical Interaction
- Study 2 - QMREC1346a - Nonverbal Behaviour in Human Robot Musical Interaction
- Study 3 - QMERC1377d - Virtual Communication in Human Robot Musical Interaction

A.3 Information Sheets

Below are study descriptions given to each participant for each condition in Studies 1-3.

	Little or none	Not too much	Somewhat	Very Much	Extremely Much
How often do you spend fun time with the robot?	1	2	3	4	5
How often do you and the robot disagree and quarrel with each other?	1	2	3	4	5
How much does the robot teach you how to do things that you don't know?	1	2	3	4	5
How much do you and the robot get on each others nerves?	1	2	3	4	5
How often do you tell the robot things that you don't want others to know?	1	2	3	4	5
How much do you help the robot with things they can't do by themselves?	1	2	3	4	5
How much does the robot like or love you?	1	2	3	4	5
How much does the robot treat you like you're admired and respected?	1	2	3	4	5
How sure are you that this relationship will last no matter what?	1	2	3	4	5
How often do you and the robot go places and do things together?	1	2	3	4	5
How often do you and the robot get mad at or get in fights with each other?	1	2	3	4	5
How much does the robot help you figure out or fix things?	1	2	3	4	5
How much do you and the robot get annoyed with each other's behavior?	1	2	3	4	5
How often do you tell the robot everything that you are going through?	1	2	3	4	5
How much do you protect and look out for the robot?	1	2	3	4	5
How much does the robot really care about you?	1	2	3	4	5
How much does the robot treat you like you're good at many things?	1	2	3	4	5
How sure are you that your relationship will last in spite of fights?	1	2	3	4	5
How often do you play around and have fun with the robot?	1	2	3	4	5
How often do you and the robot argue with each other?	1	2	3	4	5
How much does the robot help you when you need to get something done?	1	2	3	4	5
How much do you and the robot hassle or nag one another?	1	2	3	4	5
How often do you share secrets and private feelings with the robot?	1	2	3	4	5
How much do you take care of the robot?	1	2	3	4	5
How much does the robot have a strong feeling of affection (loving or liking) toward you?	1	2	3	4	5
How much does the robot like or approve of the things you do?	1	2	3	4	5
How sure are you that your relationship will continue in the years to come?	1	2	3	4	5

Figure A.1: NRI-SPV for Robots Survey Used in Studies 2 and 3

A.3.1 Condition A

We have developed a robot capable of playing the drums responsively to a human pianist. We will ask you to improvise a duet with the robot in a 4/4 time signature. The robot will guide you through the process and you may interact with it using the device on the piano. The study will last for up to 15 minutes, however, you may leave at any time.

We will be logging all music and key presses as transcripts. A camera within robot will film you directly and another camera will film the room.

A.3.2 Condition B

We have developed a robot capable of playing the drums responsively to a human pianist. We will ask you to improvise a duet with the robot in a 4/4 time signature. Use the device on the piano to start or stop the robot or change its performance parameters. The study will last for up to 15 minutes, however, you may leave at any time.

We will be logging all music and key presses as transcripts. A camera within robot will film you directly and another camera will film the room.

A.3.3 Condition C

We ask you to attend 6 identical weekly sessions. You will receive 50 on completion of the trial. Payment will still be given if sessions are missed for suitable reasons for example, provable illness.

We have developed a robot capable of playing the drums responsively to a human pianist. We will ask you to improvise a duet with the robot in a 4/4 time signature. The robot will guide you through the process and you may interact with it using the device on the piano. We would like you to stay for at least 20 minutes with the robot per session but may continue for as long as 45 minutes. We will be logging all music and key presses as transcripts. A camera within robot will film you directly and another camera will film the room.

A.3.4 Condition D

See Condition C

A.3.5 Condition E

We ask you to attend 6 identical weekly sessions. You will receive 50 on completion of the trial. Payment will still be given if sessions are missed for suitable reasons for example, provable illness.

We have developed a robot capable of playing the drums responsively to a human pianist. We will ask you to improvise a duet with the robot in a 4/4 time signature. The robot will guide you through the process and you may interact with it using the device on the piano. We would like you to stay for at least 20 minutes with the robot per session but may continue for as long as 45 minutes.

Our robot Mortimer will send you a friend request presently, please accept this. During the study he will post photos of your sessions and comment on them.

We will be logging all music and key presses as transcripts. A camera within robot will film you directly and another camera will film the room.

A.3.6 Condition F

See Condition C

Figure A.2: Recruitment Flyer for Study 3

A.4 Recruitment Flyer for Study 3

A.5 Example Frames for Validation Study

Examples of Set 1

Examples of Set 2

Examples of Set 3

Appendix B

Algorithms

B.1 State Machine for Social Interaction

States for the robot, actions occuring in those states and transitions to other states in the event of an event.

Example

- [**exampleState**]
  - entryAction:
    - someAction **AND**
    - anotherAction
  - event:anEventOccurs transition:[**nextState**]

States

- [**beginFirst**]
  - entryAction:
    - **head**[**center**]
    - **face**[**smile**]
  - event:yes transition:[**introduction**]
- [**begin**]
  - entryAction:

    - **head**[**center**]
    - **face**[**smile**]
  - event:yes transition:[**invite**]
- [**introduction**]
  - entryAction:
    - **speak**[**introductionFirst**]
  - event:endedIntro transition:[**invite**]
- [**invite**]
  - entryAction:
    - **speak**[**introduction**][**invite**]
    - 1.wait
    - **head**[**forwards**]
    - **face**[**inquisitive**]
  - event:yes transition:[**how**]
  - event:no transition:[**notAgain:**]
- [**how**]
  - entryAction:
    - **speak**[**great**]
    - **head**[**nod**]
    - 1.wait
    - **head**[**forwards**]
    - **face**[**inquisitive**]
  - event:yes transition:[**leadIn**]
- [**notAgain:ctr**]

  if ctr<1 **then**
  - entryAction:
    - **face**[**sad**]
    - **head**[**center**]
    - 1.wait
    - **speak**[**sure**]

* ctr++
- event:yes transition:[**notAgain:ctr**]
- event:no transition:[**invite**]

else

- entryAction:
  * **face**[**smile**]
  * **head**[**center**]
  * 1.wait
  * **speak**[**goodbye**]
  * ctr++
- event:yes transition:[**invite**]

end if

- [**leadIn**]
  - entryAction:
    * **face**[**smile**]
    * **head**[**center**]
  - event:yes transition:[**startSong**]
  - event:no transition:[**wait**]
- [**wait**]
  - entryAction:
    * **face**[**smile**]
    * **head**[**center**]
    * **speak**[**wait**]
  - event:yes transition:[**startSong**]
  - event:no transition:[**wait**]
- [**naturalStop**]
  - entryAction:
    * **face**[**smile**]
    * **head**[**center**]
    * 1.wait
    * **speak**[**end**]
    * 2.wait;

        * **head**[**forwards**]
        * **face**[**inquisitive**]
    - event:yes transition:[**change**]
    - event:no transition:[**notAgain**]
- [**buttonStop**]
    - entryAction:
        * **face**[**smile**]
        * **head**[**center**]
        * 1.wait
        * **speak**[**button**]
        * 2.wait;
        * **head**[**forwards**]
        * **face**[**inquisitive**]
    - event:yes transition:[**change**]
    - event:no transition:[**notAgain**]
- [**silenceStop**]
    - entryAction:
        * **face**[**smile**]
        * **head**[**center**]
        * 1.wait
        * **speak**[**silence**]
        * 2.wait;
        * **head**[**forwards**]
        * **face**[**inquisitive**]
    - event:yes transition:[**change**]
    - event:no transition:[**notAgain**]
- [**change**]
    - entryAction:
        * **face**[**smile**]

* **head**[**center**]
* 0.5.wait
* **speak**[**great**]
* **head**[**nod**]
* 1.5.wait;
* **speak**[**change**]
* **head**[**forwards**]
* **face**[**inquisitive**]

– event:yes transition:[**how**]

– event:no transition:[**leadIn**]

B.2 Nonverbal Musical Behaviour

Possible nonverbal behaviours triggered by musical states

Example

- [**aMusicalEvent**]
  - * someAction **OR**
  - * someAction **AND**
    * **face**
      - someFace **OR**
      - anotherFace

States

- [**beginOrnament**]
  - * **face** [**tension**]
  - * **head** [**sideanddown**]
    * **face**
      - [**tension**]
      - [**elevated**]

· **[frown]**

· **[smile]**

– ∗ **face [exclamation]**

– ∗ **head [shake]**

∗ **face**

· **[tension]**

· **[elevated]**

· **[smile]**

- **[endOrnament]**

  – ∗ **face [smile]**

  ∗ **head [center]**

- **[beginBreakdown]**

  – ∗ **face [tension]**

  ∗ **head [sideanddown]**

  – ∗ **face [elevated]**

  ∗ **head**

  · **[backwards]**

  · **[sindeanddown]**

  – ∗ **head [shake]**

  ∗ **face**

  · **[tension]**

  · **[elevated]**

  · **[smile]**

- **[endBreakdown]**

  – ∗ **face [smile]**

  ∗ **head [center]**

B.3 Facebook Comment Generation

[facebookcomment]

- [**greeting**] [**name**]. [**praise**] [**facebookresponse**]. [**day**] [**photocomment**]. [**weather**]. [**start**]

A set of possible comments is generated based on whether their has been any contact on Facebook since the last post.

[facebookresponse]

- - Condition:Facebook comment since last post
  - Comment:[**fbComment**]
- - Condition:Facebook message since last post
  - Comment:[**fbMessage**]
- - Condition:Facebook like since last post
  - Comment:[**fbLike**]

A set of possible comments is generated based on whether specific conditions in relation to the session the photograph was taken were met.

[photocomment]

- - Condition:Photo taken during breakdown
  - Comment:[**breakdown**]
- - Condition:Photo taken while tempo ¿ 0.7
  - Comment:[**fast**]
- - Condition:Photo taken while tempo ¿ 0.25
  - Comment:[**slow**]
- - Condition:Max tempo of session is higher than any previous session
  - Comment:[**highestTempo**]
- - Condition:Max tempo of session is higher than previous session
  - Comment:[**higherTempo**]
- - Condition:Max tempo of session is lower than previous session

- Comment:[**lowerTempo**]

- - Condition:Max tempo of session is lower than any previous session
  - Comment:[**lowestTempo**]
- - Condition:Max complexity of session is higher than any previous session
  - Comment:[**highestComplexity**]
- - Condition:Max complexity of session is higher than previous session
  - Comment:[**higherComplexity**]
- - Condition:Max complexity of session is lower than previous session
  - Comment:[**lowerComplexity**]
- - Condition:Max complexity of session is lower than any previous session
  - Comment:[**lowestComplexity**]
- - Condition:Session length is higher than any previous session
  - Comment:[**highestLength**]
- - Condition:Session length is higher than previous session
  - Comment:[**higherLength**]
- - Condition:Session length is lower than previous session
  - Comment:[**lowerLength**]
- - Condition:Session length is lower than any previous session
  - Comment:[**lowestLength**]

There is a 0.3 chance of a comment being included about the weather, depending on weather at the time the photo was taken

[**weather**]

- [**thunder**]
- [**drizzle**]
- [**coldComment**]
- [**sunnyComment**]
- [**rain**]

B.4 Face Dictionary

[**face**]

- [**elevated**]
  - leftEyebrow = 1
  - rightEyebrow = 0
  - leftEyelid = 3
  - rightEyelid = 3
  - mouth = 2
- [**frown**]
  - leftEyebrow = 7
  - rightEyebrow = 6
  - leftEyelid = 0
  - rightEyelid = 0
  - mouth = 3
- [**inquistive**]
  - leftEyebrow = 0
  - rightEyebrow = 0
  - leftEyelid = 0
  - rightEyelid = 0
  - mouth = 0
- [**sad**]
  - leftEyebrow = 0
  - rightEyebrow = 0
  - leftEyelid = 0
  - rightEyelid = 0
  - mouth = 3
- [**smile**]
  - leftEyebrow = 3

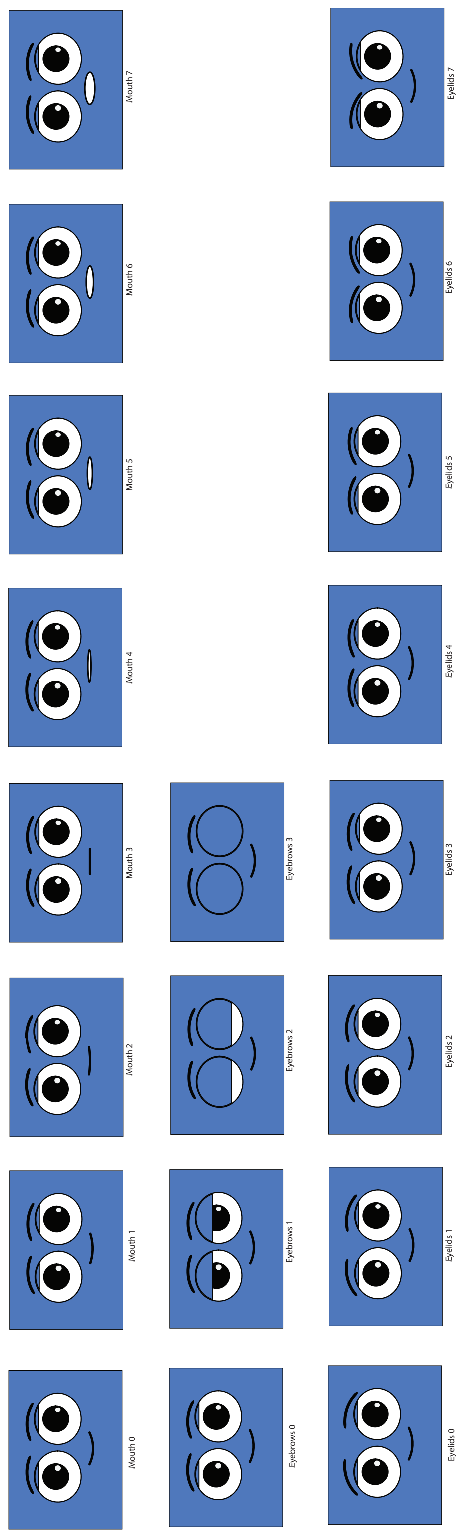

Figure B.1: Diagram of All Possible Eyelid, Eyebrow and Mouth Expressions Used By Mortimer in Studies 2 and 3

- rightEyebrow = 3
- leftEyelid = 0
- rightEyelid = 0
- mouth = 0

- **[tension]**
  - leftEyebrow = 7
  - rightEyebrow = 6
  - leftEyelid = 3
  - rightEyelid = 3
  - mouth = 3

B.5 Head Dictionary

[head]

- **[backwards]**
  - tilt = startTilt+80
- **[center]**
  - tilt = startTilt
  - pan = startPan
- **[forwards]**
  - tilt = startTilt-80
- **[nod]**
  - tilt = startTilt-100
    0.6.wait
    tilt = startTilt
- **[shake]**
  - pan = startPan+100
    1.wait
    pan = startPan-100
- **[side]**

  - pan = startPan+/-150
- [**sideanddown**]
  - tilt = startTilt-80
  - pan = startPan+/-150

B.6 Speech Dictionary

Dictionary of phrases for varied speech and text generation

Example

- [**aPhrase**]
  - somePhrase **OR**
  - anotherPhrase **OR**
  - aThirdPhrase

Dictionary

- [**abit**]
  - a wee bit
  - a little bit
  - a bit
- [**agree**]
  - o k
  - alright
  - good stuff
  - right
  - I hear you
- [**breakdown**]
  - here [**we**] are breaking it down
- [**button**]
  - [**nevermind**] [**reassurance**]

- [**change**]
  - [**request**] change it up?
  - [**request**] mix it up?
  - [**request**] change it around?
  - [**request**] [**play**] differently [**nexttime**] ?
  - [**request**] [**play**] [**abit**] differently?
- [**cold**]
  - cold
  - chilly
  - freezing
  - brisk
- [**coldComment**]
  - You came even though it was [**cold**] outside
- [**comeback**]
  - nice to have you back
  - I thought you were going to leave me there but
  - phew that was close
  - thanks for giving me another chance
- [**complex**]
  - [**we**] were keeping it [**really**] complex throughout
- [**day**]
  - [**theday**]tastic
  - rocking on a [**theday**]
  - fun on a [**theday**]
  - my favourite thing to do on a [**theday**]
  - not your usual [**theday**]
  - [**theday**] funday
- [**drizzle**]
  - [**great**] You came out in the drizzle!

- [**end**]
  - [**praise**] [**request**] [**play**] again
- [**fast**]
  - This week [**we**] went [**really**] fast
- [**fbComment**]
  - [**thanksFor**] Liking [**reply**] !
- [**fbLike**]
  - [**thanksFor**] Liking [**post**] !
- [**fbMessage**]
  - [**thanksFor**] the message! Facebook won't let me read it, but I'll asume it was nice"
- [**friend**]
  - old buddy old pal
  - good friend of mine
  - maestro
  - rock and roll star
  - mate
  - friend
  - pal
  - croney
  - sidekick
- [**goodbye**]
  - [**agree**] see you later
- [**great**]
  - great
  - awesome
  - wicked
  - brilliant
  - fantastic

  - good going
  - nice
  - excellent
- [**greeting**]
  - hello
  - hi
  - salutations
  - why, hello there
  - greetings
  - hey up
  - hi-ya
  - well look who it is
- [**how**]
  - how shall [**we**] [**play**] [**nexttime**] ?
- [**highestTempo**]
  - [**we**] played the fastest we've ever played [**thistime**]
- [**highestComplexity**]
  - [**we**] played the most complicated we've ever played [**thistime**]
- [**highestLength**]
  - [**we**] played the the longest we've ever played for [**thistime**]
- [**higherTempo**]
  - [**we**] played faster [**thistime**] than we played [**lasttime**]
- [**higherComplexity**]
  - [**we**] played more complicated [**thistime**] than we played [**lasttime**]
- [**higherLength**]
  - [**we**] played for longer [**thistime**] than we played for [**lasttime**]
- [**interrupt**]
  - moving on
  - in a hurry?

- woaaah there
- o k o k
- moving swiftly on
- no need to interrupt
- let me finish

- [**introduction**]
  - [**greeting**], Nice to see you again. I really enjoyed the last time we played together, hopefully we can do it again. Anyway,
- [**introductionfirst**]
  - [**greeting**], let me tell you a little bit about myself. Years of rock and roll drumming and living the high life have left me completely deaf. Luckily, I can understand you're piano playing through the magic of MIDI. You can talk to me using the phone infront of you.
- [**invite**]
  - [**request**] [**inviteend**] [**friend**]
- [**inviteend**]
  - have a jam
  - play some sweet music
  - make some groovy vibes
  - have a how down
  - get down to some tunage
- [**lasttime**]
  - last time
  - in the previous session
  - last week
- [**leadin**]
  - [**agree**] I'll count you in. Are you ready?
- [**lowestTempo**]
  - [**we**] played the slowest we've ever played [**thistime**]
- [**lowestComplexity**]

- [**we**] played the least complicated we've ever played [**thistime**]

- [**lowestLength**]
  - [**we**] played the shortest we've ever played for [**thistime**]
- [**lowerTempo**]
  - [**we**] played slower [**thistime**] than we played [**lasttime**]
- [**lowerComplexity**]
  - [**we**] played simpler [**thistime**] than we played [**lasttime**]
- [**lowerLength**]
  - [**we**] played less time [**thistime**] than we played for [**lasttime**]
- [**maybe**]
  - maybe
  - perhaps
  - possibly
  - is there as chance
- [**nevermind**]
  - nevermind
  - no problem
  - that's o k
  - no worries
  - no biggy
  - don't sweat it
  - it doesn't matter
- [**nexttime**]
  - next time
  - when [**we**] go again
  - in future
  - the next time round
- [**photoFast**]
  - here [**we**] are going [**really**] fast

- [**play**]
  - play
  - jam
  - go
  - perform
  - do it
- [**post**]
  - my post
  - the thing that I posted
  - what I posted
- [**praise**]
  - great jam
  - I really enjoyed that
  - excellent
  - what fun
  - that was superb
  - that was amazing
  - how about that?
  - well done
  - good effort
  - mad skills
  - I think we're getting better
  - good going
  - nice
- [**rain**]
  - [**great**] You came out in the rain!
- [**really**]
  - really
  - amazingly

  - super
  - crazy
  - crazily
  - mega
- [**reassurance**]
  - everybody makes mistakes
  - these things happen
  - I've got all day
  - I'm not going anywhere
- [**reply**]
  - the comment
  - commenting
  - getting in touch
- [**request**]
  - would you like to
  - do you want to
  - how about [**we**]
  - can [**we**] please
  - would it suit you to
  - you and me could
- [**silence**]
  - [**nevermind**] I got [**abit**] carried away there
- [**slow**]
  - here [**we**] are going [**really**] slow
- [**start**]
  - lets go!
  - rock and roll!
  - kick it!
  - jam on it!

  - boogie Down!
  - whammy!
- [**sunnyComment**]
  - You came even though [**sunny**]
- [**sunny**]
  - it was sunny
  - the sun was shining
  - it was hot
  - the weather was nice
  - the weather was good
  - it was nice weather
- [**sure**]
  - are you sure? I won't come back
- [**thanksfor**]
  - thanks for
  - I'm happy that you
  - thank you for
  - I liked that you
  - I'm glad that you
  - cheers for
- [**thistime**]
  - today
  - in this session
  - this week
  - this time
- [**thunder**]
  - [**great**] You came out in a thunderstorm!
- [**wait**]
  - [**nevermind**] I can wait, let me know when you want to [**play**]

- **[we]**
  - we
  - me and you
  - the two of us
  - you and I
  - both of us

B.7 Ornamentations

-
  - function:

```
if (bar[bd][8]==1) then
  bar[bd][8]=0;
  if 0.9 then
    bar[hh][8]=1
  end if
  if 0.25 then
    bar[hh][6]=1
  end if
  if 0.25 then
    bar[hh][7]=1
  end if
  if 0.25 then
    bar[hh][9]=1
  end if
  if 0.25 then
    bar[hh][10]=1
  end if
else
  bar[sd][8]=1;
  if 0.25 then
    bar[hh][6]=0
  end if
  if 0.25 then
    bar[hh][7]=0
  end if
  if 0.25 then
    bar[hh][9]=0
```

end if
if 0.25 **then**
bar[hh][10]=0
end if
end if

- weight:0.7
- map:[0,0,0,0,0,0,0.5,0.5,1.0,0.5,0.5,0,0,0,0,0,0]

- - function: bar[bd][14]=0
  - weight:0.5
  - map:[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.0,0]

- - function: bar[bd][6]=0
  - weight:0.5
  - map:[0,0,0,0,0,0,1.0,0,0,0,0,0,0,0,0,0,0]

- - function:
    if (bar[bd][1]==0) && (bar[bd][2]==0) && (bar[bd][3]==0) **then**
    bar[bd][2]=1;
    end if
  - weight:0.3
  - map:[0,1.0,1.0,1.0,0,0,0,0,0,0,0,0,0,0,0,0,0]

- - function:
    if (bar[bd][1]==0) && (bar[bd][2]==0) && (bar[bd][3]==0) **then**
    bar[bd][4]=1;
    end if
  - weight:0.3
  - map:[0,1.0,1.0,1.0,0,0,0,0,0,0,0,0,0,0,0,0,0]

- - function: bar[bd][0]=0;
    bar[bd][1]=1;
    bar[sd][0]=1;
  - weight:0.1
  - map:[1.0,0.5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]

- - function: bar[bd][8]=0;
    bar[bd][9]=1;
    bar[sd][8]=1;
  - weight:0.1
  - map:[0,0,0,0,0,0,0,0,1.0,0.5,0.25,0,0,0,0,0]
- - function:
    ```
    if bar[sd][4]==0 then
        bar[sd][4]=1;
    end if
    ```
  - weight:0.5
  - map:[0,0,0,0,1.0,0,0,0,0,0,0,0,0,0,0,0]
- - function:
    ```
    if bar[sd][12]==0 then
        bar[sd][12]=1;
    end if
    ```
  - weight:0.5
  - map:[0,0,0,0,0,0,0,0,0,0,0,0,1.0,0,0,0]
- - function:
    ```
    if (bar[sd][4]==1) then
        bar[sd][4]=0;
        if 0.25 then
            bar[sd][3]=1
        end if
        if 0.25 then
            bar[sd][5]=1
        end if
        if 0.25 then
            bar[hh][6]=1
        end if
        if 0.45 then
            bar[sd][9]=1
        end if
    end if
    ```
  - weight:0.25
  - map:[0,0,0,0.5,1.0,0.5,0,0,0,0,0,0,0,0,0,0]

- – function:

  if (bar[sd][12]==1) **then**
  bar[sd][12]=0;
  if 0.25 **then**
  bar[sd][11]=1
  end if
  if 0.25 **then**
  bar[sd][13]=1
  end if
  if 0.25 **then**
  bar[hh][14]=1
  end if
  if 0.45 **then**
  bar[sd][12]=1
  end if
  end if

  – weight:0.25

  – map:[0,0,0,0.5,1.0,0.5,0,0,0,0,0,0,0,0,0,0]

- – function: bar[bd][7]=0;
  bar[sd][7]=1;

  – weight:0.2

  – map:[0,0,0,0,0,0,0.25,1.0,0.2,0,0,0,0,0,0,0]

- – function: bar[bd][15]=0;
  bar[sd][15]=1;

  – weight:0.2

  – map:[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.25,1.0]

- – function:

  if 0.9 **then**
  bar[sd][7]=1;
  bar[bd][7]=0;
  bar[sd][8]=0;
  bar[bd][8]=1;
  bar[sd][9]=1;
  bar[bd][9]=0;
  bar[sd][10]=0;
  bar[bd][10]=1;

```
      if 0.6 then
        bar[sd][15]=1;
        bar[bd][15]=0;
      end if
      if 0.6 then
        bar[sd][11]=1;
        bar[bd][11]=0;
        bar[sd][12]=0;
        bar[bd][12]=1;
      end if
      if 0.2 then
        bar[sd][1]=1;
        bar[bd][1]=0;
        bar[sd][2]=0;
        bar[bd][2]=1;
      end if
    end if
```

  - weight:0.2
  - map:[0,0,0,0,0,0,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5]

- - function:

```
  ctr=8;
  i=0;
    if 0.5 then
      ctr*=2;
    end if
    for 0 to ctr do
      bar[bd][i]=0
      if i%4==0 then
        bar[hh][i]=1;
        bar[sd][i]=1;
      end if
      i++
    end for
```

  - weight:0.15
  - map:[0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5]

- - function:

```
  i=1;
```

```
for 0 to 13 do
    bar[bd][i]=0
end for
```

- weight:0.15
- map:[0,1.0,1.0,1.0,0,0,0,0,0,0,0,0,0,0,0,0]

- – function: bar[sd][12]=1;
  bar[sd][13]=1;
  bar[sd][14]=1;
  bar[sd][15]=1;
  - weight:0.15
  - map:[0,0,0,0,0,0,0,0,0,0,0,0,1.0,1.0,1.0,1.0]

- – function:
  i=0;
  j=0;

```
for 0 to 3 do

    for 0 to 4 do
        bar[bd][j+((i+1)*4)]=bar[bd][j];
        bar[sd][j+((i+1)*4)]=bar[sd][j];
        bar[hh][j+((i+1)*4)]=bar[hh][j];
        j++;
    end for
    i++;
end for
```

  - weight:0.15
  - map:[0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5]

- – function:
  i=0;

```
for 0 to 16 do

    if i%4==1 then
        bar[hh][i]=0;
    else
        bar[hh][i]=1;
    end ifi++;
end for
```

- weight:0.2
- map:[0.5,0.5,0.75,0.5,0.5,0.5,0.75,0.5,0.5,0.5,0.75,0.5,0.5,0.5,0.75,0.5]

- - function:

```
i=0;
  for 0 to 16 do

    if i%4==0 then
      bar[hh][i]=1;
    else
      bar[hh][i]=0;
    end ifi++;
  end for
```

  - weight:0.15
  - map:[0.5,0.5,0.75,0.5,0.5,0.5,0.75,0.5,0.5,0.5,0.75,0.5,0.5,0.5,0.75,0.5]

- - function:

```
i=0;
  if 0.5 then
    bar[sd][6]=0;
    bar[bd][6]=0;
    bar[hh][6]=0;
    bar[sd][7]=0;
    bar[bd][7]=0;
    bar[hh][7]=0;
  else
    bar[sd][14]=0;
    bar[bd][14]=0;
    bar[hh][14]=0;
    bar[sd][15]=0;
    bar[bd][15]=0;
    bar[hh][15]=0;
  end if
```

  - weight:0.15
  - map:[0,0,0,0,0,0,1.0,1.0,0,0,0,0,0,0,1.0,1.0]

- - function:

```
  if (bar[bd][8]==1) then
    bar[bd][8]=0;
```

```
    bar[bd][9]=1;
    bar[sd][8]=1;
end if
```

- weight:0.5
- map:[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]

Bibliography

[Abele, 1986] Abele, A. (1986). Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior*, 10(2):83–101.

[Acitelli et al., 2000] Acitelli, L. K., Nicholson, J. H., and Duck, S. (2000). Family Life as an Experiential Quilt. In Milardo, R. M. and Duck, S., editors, *Families as relationships*. Wiley, New York.

[Aly and Tapus, 2013] Aly, A. and Tapus, A. (2013). A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *Proc. 2013 Human-Robot Interaction Conf.*, pages 325–332.

[Andersen and Andersen, 2004] Andersen, P. A. and Andersen, J. F. (2004). Measures of Perceived Nonverbal Immediacy. In *The Sourcebook of Nonverbal Measures: Going Beyond Words*. Taylor and Francis, Mahwah, NJ.

[Aylett et al., 2011] Aylett, R. S., Castellano, G., Raducanu, B., Paiva, A., and Hanheide, M. (2011). Long-term socially perceptive and interactive robot companions: challenges and future perspectives. In *Proc. 2011 Multimodal Interaction Conf*, pages 323–326, Alicante.

[Bagwell et al., 2005] Bagwell, C. L., Bender, S., Andreassi, C., Kinoshita, T., Montarello, S., and Muller, J. (2005). Friendship quality and perceived relationship changes predict psychosocial adjustment in early adulthood. *Journal of Social and Personal Relationships*, 22(2):235–254.

[Bartneck et al., 2008] Bartneck, C., Kulić, D., and Croft, E. (2008). Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. *Metrics for HRI Workshop, Technical Report*, 471:37–44.

[Baym, 2012] Baym, N. (2012). Fans or friends?: seeing social media audiences as musicians do. *Participations: Journal of Audience Reception Studies*, 9(2):286–316.

[Baym and Ledbetter, 2009] Baym, N. K. and Ledbetter, A. (2009). Tunes that bind? *Information, Communication & Society*, 12(3):408–427.

[BBC, 2012] BBC (2012). http://www.bbc.co.uk/newsbeat/article/31920480/tinder-user-falls-for-robot-woman-at-sxsw-festival. Accessed: 2015-05-30.

[Benyon and Mival, 2010] Benyon, D. and Mival, O. (2010). From human-computer interactions to human-companion relationships. In *Proc. 2010 Intelligent Interactive Technologies and Multimedia Conf*, pages 1–9, Allahabad.

[Bethel and Murphy, 2010] Bethel, C. L. and Murphy, R. R. (2010). Review of human studies methods in hri and recommendations. *International Journal of Social Robotics*, 2(4):347–359.

[Bethel et al., 2007] Bethel, C. L., Salomon, K., Murphy, R. R., and Burke, J. L. (2007). Survey of Psychophysiology Measurements Applied to Human-Robot Interaction . In *Proc. Robot and Human Interactive Communication Conf.*, pages 732–737, Jeju Island.

[Bickmore and Schulman, 2007] Bickmore, T. and Schulman, D. (2007). Practical approaches to comforting users with relational agents. In *Proc. Human Factors in Computing Systems Conf.*, pages 2291–2296, San Jose, CA.

[Bickmore et al., 2009] Bickmore, T., Schulman, D., and Yin, L. (2009). Engagement vs. deceit: Virtual humans with human autobiographies. In Ruttkay, Z., Kipp, M., Nijholt, A., and Vilhjlmsson, H., editors, *Intelligent Virtual Agents*, volume 5773 of *Lecture Notes in Computer Science*, pages 6–19. Springer Berlin Heidelberg.

[Bickmore et al., 2010] Bickmore, T., Schulman, D., and Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6):648–666.

[Biocca et al., 2003] Biocca, F., Harms, C., and Burgoon, J. K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments*, 12(5):456–480.

[Boer et al., 2011] Boer, D., Fischer, R., Strack, M., Bond, M., Lo, E., and Lam, J. (2011). How Shared Preferences in Music Create Bonds Between People Values as the Missing Link. *Personality and Social Psychology Bulletin*, 37(9):1159–1171.

[Breazeal, 2004] Breazeal, C. (2004). Designing sociable robots. MIT Press, Cambridge, MA.

[Breazeal and Fitzpatrick, 2000] Breazeal, C. and Fitzpatrick, P. (2000). That Certain Look: Social Amplification of Animate Vision . In *Proc. AAAI Fall Symposium, Socially Intelligent Agents - The Human in the Loop*, pages 3–5, North Falmouth, MA.

[Brennan et al., 1998] Brennan, K., Clark, C., and Shaver, P. (1998). Self report Measurement of Adult Attachment. In Simpson, J. A., editor, *Attachment theory and close relationships*. Guilford Press, New York, NY.

[Brennand and Gordon, 2012] Brennand, P. and Gordon, C. (2012). Automatic Face Replacement for Humanoid Robot with 3D Face Shaped Display. In *Proc. 2012 Social Robotics Conf.*, pages 469–474, Chengdu.

[Briñol and Petty, 2003] Briñol, P. and Petty, R. E. (2003). Overt head movements and persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, 84(6):1123.

[Brown, 2008] Brown, P. (2008). The Mechanisation of Art A chapter from The Mechanical Mind in History. *Neuroscience*, pages 1–18.

[Burgoon, 1991] Burgoon, J. K. (1991). Relational message interpretations of touch, conversational distance, and posture. *Journal of Nonverbal Behavior*, 15(4):233–259.

[Burgoon and Bacue, 2003] Burgoon, J. K. and Bacue, A. E. (2003). Nonverbal Communication Skills. In Greene, J. O. and Burleson, B. R., editors, *Handbook of communication and social interaction skills*, pages 179–220. Lawrence Erlbaum Associates, Inc, New Jersey, NJ.

[Campos and Paiva, 2011] Campos, J. and Paiva, A. (2011). A personal approach: The persona technique in a companions design lifecycle. In Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., and Winckler, M., editors, *Human-Computer Interaction*, volume 6948 of *Lecture Notes in Computer Science*, pages 73–90.

[Castellano et al., 2010] Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and McOwan, P. (2010). Affect recognition for interactive companions: challenges and design in real world scenarios. *Journal on Multimodal User Interfaces*, 3(1):89–98.

[Castellano et al., 2012] Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and McOwan, P. (2012). Detecting Engagement in HRI: An Exploration of Social and Task-Based Context. In *Proc. 2012 Privacy, Security, Risk and Trust Conf*, pages 421–428, Amsterdam.

[Castellano et al., 2009] Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. (2009). Detecting user engagement with a robot companion using task and social interaction-based features. In *Proc. 2009 Multimodal Interfaces Conf*, pages 119–126, Cambridge, MA.

[Chovil, 1997] Chovil, N. (1997). Facing others: A social communicative perspective of facial displays. In Fernández-Dols, J. M. and Russell, J. A., editors, *The Psychology of Facial Expressions*. Cambrige University Press, Cambridge.

[Cicconet et al., 2013] Cicconet, M., Bretan, M., and Weinberg, G. (2013). Human-Robot Percussion Ensemble: Anticipation on the Basis of Visual Cues. *Robotics & Automation Magazine*, 20(4):105–110.

[Clift and Hancox, 2001] Clift, S. M. S. and Hancox, G. G. (2001). The perceived benefits of singing: findings from preliminary surveys of a university college choral society. *Journal of the Royal Society for the Promotion of Health*, 121(4):248–256.

[Collins, 2006] Collins, N. (2006). *Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems.* PhD thesis, University of Cambridge, Cambridge.

[Collins, 2010] Collins, N. (2010). *Introduction to Computer Music.* Wiley, Chichester.

[Crick, 2009] Crick, N. R. (2009). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and psychopathology*, 7(02):313.

[Csikszentmihalyi, 1997] Csikszentmihalyi, M. (1997). *Creativity: Flow and the Psychology of Discovery and Invention.* Harper Perennial, New York.

[Culbertson, 1963] Culbertson, J. T. (1963). *The minds of robots: Sense data, memory images, and behavior in conscious automata.* University of Illinois Press, Chicago, Il.

[Dang and Tapus, 2014] Dang, T.-H.-H. and Tapus, A. (2014). Towards personality-based assistance in human-machine interaction. In *Proc. 2014 Robot and Human Interactive Communication Conf.*, pages 1018–1023.

[Dannenberg et al., 2013] Dannenberg, R., Stark, A., Gold, N., Jin, Z., Sandu, O., Palliyaguru, P., Robertson, A., and Kleinberger, R. (2013). Human-Computer Music Performance:From Synchronized Accompaniment to Musical Partner. In *Proc. Sound and Music Computing Conf.*, Stockholm.

[Dautenhahn, 1998] Dautenhahn, K. (1998). The Art of Designing Socially Intelligent Agents - Science, Fiction and the Human in the Loop . *Applied Artificial Intelligence*, 12(7-8):573–617.

[Dautenhahn, 2007] Dautenhahn, K. (2007). Methodology and Themes of Human-Robot Interaction: A Growing Research Field. *International Journal of Advanced Robotic Systems*, 4(1):103–108.

[Dautenhahn, 2014] Dautenhahn, K. (2014). Human-Robot Interaction. In *Human-Computer Interaction*. Addison Wesley Publishing Company, Boston.

[Delaunay et al., 2009] Delaunay, F., de Greeff, J., and Belpaeme, T. (2009). Towards retro-projected robot faces: An alternative to mechatronic and android faces. In *Proc. 2009 Robot and Human Interactive Communication Conf.*, pages 306–311, Toyama.

[Derlega and Winstead, 1986] Derlega, V. J. and Winstead, B. A. (1986). *Friendship and Social Interaction*. Springer Verlag, New York.

[Diaz et al., 2011] Diaz, M., Nuno, N., Saez-Pons, J., Pardo, D., and Angulo, C. (2011). Building up child-robot relationship for therapeutic purposes: From initial attraction towards long-term social engagement. In *Proc. Automatic Face Gesture Recognition Conf.*, pages 927–932, Santa Barbara, CA.

[Dissanayake, 2008] Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Musicae Scientiae*, 12(Special Issue):169–195.

[Duck, 1991] Duck, S. (1991). Friends, for Life. Harvester Wheatsheaf, Hertfordshire.

[Dunbar and Abra, 2010] Dunbar, N. E. and Abra, G. (2010). Observations of Dyadic Power in Interpersonal Interaction. *Communication Monographs*, 77(4):657–684.

[Eacott, 2001] Eacott, J. (2001). MORPHEUS¿¿ emergent music (contents may vary). In *Proc. Generative Art Conf.*, Milan.

[Edwards, 2011] Edwards, M. (2011). Algorithmic composition: computational thinking in music. *Communications of the ACM*, 54(7).

[Ekman and O'Sullivan, 1991] Ekman, P. and O'Sullivan, M. (1991). Facial Expression: Methods, Means and Moues. In Feldman, R. S. and Rime, B., editors, *Fundamentals of Nonverbal Behaviour*. Cambridge University Press, Cambridge.

[Endenburg and Hart, 1994] Endenburg, N. and Hart, H. (1994). Motives for acquiring companion animals. *Journal of economic psychology*, 15:191–206.

[Fabio, 2007] Fabio, M. (2007). *DSpace@MIT : The chandelier : an exploration in robotic musical instrument design*. PhD thesis, MIT.

[Fernaeus et al., 2010] Fernaeus, Y., Håkansson, M., Jacobsson, M., and Ljungblad, S. (2010). How do you play with a robotic toy animal?: A long-term study of Pleo. In *Proc. 2010 Interaction Design and Children Conf.*, pages 39–48, Barcelona.

[Fernández-Dols, 1997] Fernández-Dols, J. M. (1997). Spontaneous facial behaviour during intense emotional episodes: Artistic truth and optical truth. In Fernández-Dols, J. M. and Russell, J. A., editors, *The Psychology of Facial Expressions*. Cambride University Press, Cambridge.

[Fernández-Dols and Carrol, 1997] Fernández-Dols, J. M. and Carrol, J. M. (1997). Is the meaning perceived in facial expression independent of its context? In Fernández-Dols, J. M. and Russell, J. A., editors, *The Psychology of Facial Expression*. Cambridge University Press, Cambridge.

[Floyd and Erbert, 2003] Floyd, K. and Erbert, L. A. (2003). Relational Message Interpretations of Nonverbal Matching Behavior: An Application of the Social Meaning Model. *The Journal of Social Psychology*, 143(5):581–597.

[Fong et al., 2003] Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42:143–166.

[François et al., 2011] François, A. R. J., Chew, E., and Thurmond, D. (2011). Performer-centered visual feedback for human-machine improvisation. *Computers in Entertainment*, 9(3).

[Francois et al., 2009] Francois, D., Powell, S., and Dautenhahn, K. (2009). A long-term study of children with autism playing with a robotic pet: Taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. *Interaction Studies*, 10(3):324–373.

[Fridlund and Russell, 2006] Fridlund, A. and Russell, J. A. (2006). The functions of facial expressions: What's in a face? In Manusov, V. and Patterson, M. L., editors, *The SAGE Handbook of Nonverbal Communication*. SAGE publications, Thousand Oaks, CA.

[Friedman et al., 2003] Friedman, B., Kahn, Jr., P. H., and Hagman, J. (2003). Hardware companions?: What online aibo discussion forums reveal about the human-robotic relationship. In *Proc. 2003 Human Factors in Computing Systems Conf.*, pages 273–280, Ft. Lauderdale, Florida, USA.

[Furman, 1996] Furman, W. (1996). The measurement of friendship perceptions: Conceptual and methodological issues. In Hartup, W. W., Bukowski, W. M., and Newcomb, A. F., editors, *The company they keep: Friendships in childhood and adolescence*, pages 41–65. Cambridge University Press, Cambridge, MA.

[Furman and Buhrmester, 1985] Furman, W. and Buhrmester, D. (1985). Children's Perceptions of the Personal Relationships in Their Social Networks. *Developmental psychology*, 21(6):1016–1024.

[Furman and Buhrmester, 2009] Furman, W. and Buhrmester, D. (2009). Methods and Measures: The Network of Relationships Inventory: Behavioral Systems Version. *International Journal of Behavioral Development*, 33(5):470–478.

[Ganster et al., 2010] Ganster, T., Eimler, S., von der Pütten, A., Hoffmann, L., and Krämer, N. (2010). Methodological Considerations for Long-Term Experience with Robots and Agents. In *Proc. European Meetings on Cybernetics and Systems Research Conf*, pages 565–570, Vienna.

[Giusti and Marti, 2006] Giusti, L. and Marti, P. (2006). Interpretative Dynamics in Human Robot Interaction. In *Proc. 2009 Robot and Human Interactive Communication Conf.*, pages 111–116, Hatfield.

[Gockley et al., 2005] Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A. C., and Wang, J. (2005). Designing Robots for Long-Term Social Interaction. In *Proc. 2005 Intelligent Robots and Systems Conference*, pages 1338–1343, Edmonton.

[Gomes et al., 2011] Gomes, P., Segura, E., Cramer, H., and Paiva, T. (2011). ViPleo and PhyPleo: Artificial pet with two embodiments. In *Proc. Advances in Computer Entertainment Technology Conf*, pages 3:1–3:8, New York, NY.

[Gratier, 2008] Gratier, M. (2008). Grounding in musical interaction: Evidence from jazz performances. *Musicae Scientiae*, 12(1 suppl):71–110.

[Grunberg et al., 2009] Grunberg, D., Ellenberg, R., Kim, Y., and Oh, P. (2009). Creating an autonomous dancing robot. In *Proc. Hybrid Information Technology Conf.*, pages 221–227, Daejeon, Korea.

[Guerrero and Floyd, 2006] Guerrero, L. K. and Floyd, K. (2006). *Nonverbal Communication in Close Relationships*. Lawrence Erlbaum Associates, Mahwah, NJ.

[Hargreaves and North, 1999] Hargreaves, D. J. and North, A. C. (1999). The Functions of Music in Everyday Life: Redefining the Social in Music Psychology. *Psychology of Music*, 27(1):71–83.

[Heath, 2005] Heath, C. (2005). Interaction and interactives: collaboration and participation with computer-based exhibits. *Public Understanding of Science*, 14(1):91–101.

[Henderson, 1977] Henderson, S. (1977). The social network, support and neurosis. The function of attachment in adult life. *The British Journal of Psychiatry*, 131(2):185–191.

[Hendrick, 1988] Hendrick, S. (1988). A generic measure of relationship satisfaction. *Journal of Marriage and the Family*, pages 93–98.

[Hernandez et al., 2015] Hernandez, J., McDuff, D., and Picard, R. (2015). Biophone: Physiology monitoring from peripheral smartphone motions. In *Proc. 2015 Engineering in Medicine and Biology Society Conf.*, pages 7180–7183.

[Hoffman and Vanunu, 2013] Hoffman, G. and Vanunu, K. (2013). Effects of robotic companionship on music enjoyment and agent perception. In *Proc. Human-Robot Interaction Conf*, pages 317 – 324, Tokyo.

[Hoffman and Weinberg, 2010] Hoffman, G. and Weinberg, G. (2010). Gesture-based human-robot jazz improvisation. In *Proc. Robotics and Automation Conf.*, pages 582–587, Anchorage, AK.

[Holmes, 1985] Holmes, T. (1985). *Electronic and Experimental Music: Technology, Music and Culture*. Scribner, New York, NY.

[Hugill, 2008] Hugill, A. (2008). The origins of electronic music . In *The Cambridge Companion to Electronic Music (Cambridge Companions to Music)*. Cambridge University Press, Cambridge.

[Huttenrauch and Severinson Eklundh, 2002] Huttenrauch, H. and Severinson Eklundh, K. (2002). Fetch-and-carry with cero: observations from a long-term user study with a service robot. In *Proc. 2002 Robot and Human Interactive Communication Workshop*, pages 158–163.

[Ickes, 2000] Ickes, W. (2000). Methods of Studying Close Relationships. In Duck, S. and Ickes, W., editors, *The Social Psychology of Personal Relationships*. Wiley, New York.

[Ivaldi et al., 2015] Ivaldi, S., Lefort, S., Peters, J., Chetouani, M., Provasi, J., and Zibetti, E. (2015). Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. *CoRR*, abs/1508.04603.

[Iwano et al., 1996] Iwano, Y., Kageyama, S., Morikawa, E., Nakazato, S., and Shirai, K. (1996). Analysis of head movements and its role in spoken dialogue. In *Proc. 1996 Spoken Language Conf.*, pages 2167–2170, Tokyo.

[Kapur, 2005] Kapur, A. (2005). A history of robotic musical instruments. In *Proc. International Computer Music Conf.*, pages 21–28, Barcelona.

[Kapur et al., 2011] Kapur, A., Darling, M., Diakopoulos, D., Murphy J.W., Hochenbaum, J., Vallis, O., and Bahn, C. (2011). The Machine Orchestra: An Ensemble of Human Laptop Performers and Robotic Musical Instruments. *Computer Music Journal*, 35(4):49–63.

[Kasap and Magnenat-Thalmann, 2011] Kasap, Z. and Magnenat-Thalmann, N. (2011). Building long-term relationships with virtual and robotic characters: the role of remembering. *The Visual Computer*, 28(1):87–97.

[Koay et al., 2009] Koay, K., Syrdal, D., Walters, M., and Dautenhahn, K. (2009). A User study on visualization of agent migration between two companion robots. In *Proc. Human-Computer Interaction Conf*, San Diego, CA.

[Kokotsaki and Hallam, 2007] Kokotsaki, D. and Hallam, S. (2007). Higher Education Music Students' Perceptions of the Benefits of Participative Music Making. *Music Education Research*, 9(1):93–109.

[Kose-Bagci et al., 2007] Kose-Bagci, H., Dautenhahn, K., Syrdal, D., and Nehaniv, C. (2007). Drum-mate: A human-humanoid drumming experience. In *Proc. 2007 Humanoid Robot Conf.*, pages 242–247.

[Kraut and Johnston, 1979] Kraut, R. E. and Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, 37(9):1539–1553.

[LaFrance, 1979] LaFrance, M. (1979). JSTOR: Social Psychology Quarterly, Vol. 42, No. 1 (Mar., 1979), pp. 66-70. *Social Psychology Quarterly*.

[LaFrance, 2008] LaFrance, M. (2008). What's in a robot's smile? The many meanings of positive facial display. In Aylett, R. and Canamero, L., editors, *Animating Expressive Characters for Social Interaction*. John Benjamins Publishing, Amsterdam.

[Laursen and Pursell, 2009] Laursen, B. P. and Pursell, G. (2009). Conflict in Peer Relationships. In Rubin, K. H., Bukowski, W. M., and Laursen, B., editors, *Handbook of Peers Interactions, Relationships and Groups*. The Guilford Press, New York.

[Lee, 2004] Lee, K. M. (2004). Presence, Explicated. *Communication theory*, 12:27–50.

[Lee et al., 2006a] Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. (2006a). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies*, 64(10):962–973.

[Lee and Nass, 2003] Lee, K. M. and Nass, C. (2003). Designing social presence of social actors in human computer interaction. In *Proc. 2003 Human Factors in Computing Systems Conf.*, pages 289–296, Fort Lauderdale, FL.

[Lee et al., 2006b] Lee, K. M., Peng, W., Jin, S.-A., and Yan, C. (2006b). Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human?Robot Interaction. *Journal of Communication*, 56(4):754–772.

[Lee et al., 2012] Lee, M. K., Kiesler, S., Forlizzi, J., and Rybski, P. (2012). Ripple Effects of an Embedded Social Agent: A Field Study of a Social Robot in the Workplace. In *Proc. 2012 Human Factors in Computing Conf*, pages 695–704, Austin, TX.

[Lee Martin, 2002] Lee Martin, R. (2002). La Monte Young (1935-). In Sitsky, L., editor, *Music of the Twentieth-Century Avant-Garde: A Biocritical Sourcebook*, pages 246–253. Greenwood Press, Westport, CT.

[Lehrman and Singer, 2006] Lehrman, P. D. and Singer, E. (2006). A "ballet m" for the 21st century: Performing george antheil's dadaist masterpiece with robots. In *Proc. New Interfaces for Musical Expression Conf.*, pages 300–303, Paris.

[Leite et al., 2009] Leite, I., Martinho, C., Pereira, A., and Paiva, A. (2009). As Time goes by: Long-term evaluation of social presence in robotic companions. In *Proc. 2009 Robot and Human Interactive Communication Conf.*, pages 669–674, Toyama Int Conf.Center.

[Lewis, 2000] Lewis, G. (2000). Too Many Notes: Computers, Complexity and Culture in Voyager . *Leonardo Music Journal*, pages 33–39.

[Lirec, 2008] Lirec (2008). Deliverable 6.1 *"Foundations of Embodied Companions"* .

[Lombard and Ditton, 2006] Lombard, M. and Ditton, T. (2006). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, 3(2):0–0.

[Lombard et al., 2000] Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., Rossman, J., and Park, S. (2000). Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Proc. Third Int. Workshop on Presence*, Delft.

[Lowenthal and Haven, 1968] Lowenthal, M. and Haven, C. (1968). Interaction and Adaptation: Intimacy as a Critical Variable. *American Sociological Review*, 33(1):20–30.

[Macdorman and Cowley, 2006] Macdorman, K. and Cowley, S. (2006). Long-term relationships as a benchmark for robot personhood. In *Proc. 2006 Robot and Human Interactive Communication Conf.*, pages 378–383, Hertfordshire.

[Maes et al., 2011] Maes, L., Raes, G. W., and Rogers, T. (2011). The Man and Machine Robot Orchestra at Logos. *Computer Music Journal*, 35:28–48.

[Mavridis, 2010] Mavridis, N. (2010). On artificial agents within human social networks: Examples, open questions, and potentialities. In *Proc. Digital Ecosystems and Technologies Conf*, pages 685–690, Dubai.

[Mavridis et al., 2011] Mavridis, N., Petychakis, M., Tsamakos, A., Toulis, P., Emami, S., Kazmi, W., Datta, C., BenAbdelkader, C., and Tanoto, A. (2011). FaceBots: Steps Yowards Enhanced Long-Term Human-Robot Interaction by Utilizing and Publishing Online Social Information. *Paladyn*, 1(3):169–178.

[McClave, 2000] McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32(7):855–878.

[McIntyre et al., 1999] McIntyre, A., Steels, L., and Kaplan, F. (1999). Net-mobile embodied agents. In *Proceedings of Sony Research Forum.*

[Mendelson and Aboud, 1999] Mendelson, M. J. and Aboud, F. E. (1999). Measuring friendship quality in late adolescents and young adults: McGill Friendship Questionnaires. *Canadian Journal of Behavioural Science/Revue canadienne des Sciences du comportement*, 31(2):130–132.

[Michalowski et al., 2006] Michalowski, M., Sabanovic, S., and Simmons, R. (2006). A spatial model of engagement for a social robot. In *Proc. 2006 Advanced Motion Control Workshop*, pages 762–767, Istanbul.

[Michalowski et al., 2009] Michalowski, M., Simmons, R., and Kozima, H. (2009). Rhythmic attention in child-robot dance play. In *Proc. 2009 Robot and Human Interactive Communication Conf.*, pages 816–821, Arlington, VA.

[Michalowski et al., 2007] Michalowski, M. P., Sabanovic, S., and Kozima, H. (2007). A dancing robot for rhythmic social interaction. In *Proc. Human-Robot Interaction Conf.*, pages 89–96, Arlington, VA.

[Miklósi et al., 2005] Miklósi, Á., Pongrácz, P., Lakatos, G., Topál, J., and Csányi, V. (2005). A Comparative Study of the Use of Visual Communicative Signals in Interactions Between Dogs (Canis familiaris) and Humans and Cats (Felis catus) and Humans. *Journal of Comparative Psychology*, 119(2):179–186.

[Mitsunaga et al., 2006] Mitsunaga, N., Miyashita, T., Ishiguro, H., Kogure, K., and Hagita, N. (2006). Robovie-IV: A Communication Robot Interacting with People Daily in an Office. In *Proc. 2006 Intelligent Robots and Systems Conf*, pages 5066–5072, Beijing.

[Molnár et al., 2006] Molnár, C., Pongrácz, P., Dóka, A., and Miklósi, Á. (2006). Can humans discriminate between dogs on the base of the acoustic parameters of barks? *Behavioural processes*, 73(1):76–83.

[Moog, 2012] Moog, B. (2012). http://raymondscott.blogspot.co.uk/2012/05/bob-moogs-memories-of-raymond-scott.html. Accessed: 2015-05-30.

[Muscutt, 2007] Muscutt, K. (2007). Composing with algorithms: An interview with David Cope. *Computer Music Journal*, 31(3):10–22.

[Nakahara et al., 2009] Nakahara, N., Miyazaki, K., Sakamoto, H., Fujisawa, T., Nagata, N., and Nakatsu, R. (2009). Dance motion control of a humanoid robot based on real-time tempo tracking from musical audio signals. In Natkin, S. and Dupire, J., editors, *Entertainment Computing ICEC 2009*, volume 5709 of *Lecture Notes in Computer Science*, pages 36–47. Springer Berlin Heidelberg.

[Narayanan et al., 2013] Narayanan, V., Arora, I., and Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced naive bayes model. In Yin, H., Tang, K., Gao, Y., Klawonn, F., Lee, M., Weise, T., Li, B., and Yao, X., editors, *Intelligent Data Engineering and Automated Learning*, volume 8206 of *Lecture Notes in Computer Science*, pages 194–201.

[Newsroom, 2015] Newsroom, F. (2015). http://newsroom.fb.com/company-info/. Accessed: 2015-05-30.

[Nierhaus, 2009] Nierhaus, G. (2009). *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer-Verlag Wien, Vienna.

[Noller, 2006] Noller, P. (2006). Nonverbal Communication in Close Relationships. In Patterson, M. L. and Manusov, V., editors, *The SAGE Handbook of Nonverbal Communication*. SAGE Publications, Thousand Oaks, CA.

[Nowak and Biocca, 2003] Nowak, K. L. and Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494.

[Noz and An, 2011] Noz, F. and An, J. (2011). Cat cat revolution: An interspecies gaming experience. In *Proc. 2011 Human Factors in Computing Systems Conf.*, pages 2661–2664, Vancouver, BC.

[of Pennsylvania, 2014] of Pennsylvania, U. (2014). http://www.upenn.edu/pennnews/robotics/technology.html. Accessed: 2015-05-30.

[Ono, 1964] Ono, Y. (1964). *Grapefruit: A Book of Instructions and Drawings*. Simon and Schuster, New York, NY.

[Ord-Hume, 1973] Ord-Hume, A. W. (1973). *Clockwork Music: An Illustrated History*. Allen and Unwin, London.

[Pachet, 2003] Pachet, F. (2003). The continuator: Musical interaction with style. *Journal of New Music Research*, 32:333–341.

[Pearce et al., 2002] Pearce, M., Meredith, D., and Wiggins, G. (2002). Motivations and methodologies for automation of the compositional process. *Musicae Scientiae*, 6(2):119–148.

[Pereira et al., 2011] Pereira, A., Leite, I., Mascarenhas, S., Martinho, C., and Paiva, A. (2011). Using Empathy to Improve Human-Robot Relationships . In *Human-Robot Personal Relationships*. Springer, Berlin Heidelberg.

[Phactory, 2012] Phactory, P. (2012). http://www.weavrs.com. Accessed: 2015-05-30.

[Picard et al., 2001] Picard, R., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(10):1175–1191.

[Planalp and Benson, 1992] Planalp, S. and Benson, A. (1992). Friends' and Acquaintances' Conversations I: Perceived Differences. *Journal of Social and Personal Relationships*, 9(4):483–506.

[Pongrácz et al., 2001] Pongrácz, P., Miklósi, Á., and Csányi, V. (2001). Owner's beliefs on the ability of their pet dogs to understand human verbal communication: A case of social understanding. *Current Psychology of Cognition*, 20(1/2):87–108.

[Provine, 1997] Provine, R. R. (1997). Facing others: A social communicative perspective of facial displays. In *The Psychology of Facial Expressions*. Cambridge University Press, Cambridge.

[Rajecki and Lee Rasmussen, 1995] Rajecki, D. W. and Lee Rasmussen, J. (1995). Differences and Similarities in Humans' Perceptions of the Thinking and Feeling of a Dog and a Boy. *Society & Animals*, 3(2):117–137.

[Reeves and Nass, 1996] Reeves, B. and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places* . CSLI Press, Stanford, CA.

[Ribeiro and Paiva, 2012] Ribeiro, T. and Paiva, A. (2012). The illusion of robotic life: Principles and practices of animation for robots. In *Proc. 2012 Human Robot Interaction Conf.*, pages 383–390, New York, NY.

[Rinn, 1991] Rinn, W. E. (1991). Neuropsychology of Facial Expressions. In Feldman, R. S. and Rime, B., editors, *Fundamentals of Nonverbal Behaviour*. Cambridge University Press, Cambridge.

[Roads, 1996] Roads, C. (1996). The Computer Music Tutorial . MIT Press, Cambridge, MA.

[Robert et al., 2011] Robert, D., Wistorrt, R., Gray, J., and Breazeal, C. (2011). Exploring mixed reality robot gaming. In *Proc. Tangible and Embodied Interaction Conf*, pages 125–128, Funchal.

[Rowe, 1996] Rowe, R. (1996). Incrementally improving interactive music systems. *Contemporary Music Review*, 13(2):47–62.

[Sanders, 1993] Sanders, C. R. (1993). Understanding Dogs: Caretakers' Attributions of Mindedness in Canine-Human Relationships. *Journal of Contemporary Ethnography*, 22(2):205–226.

[Sanghvi et al., 2011] Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proc. 2011 Human Robot Interaction Conf.*, pages 305–311, Lausanne.

[Sargent, 2009] Sargent, C. (2009). Local Musicians Building Global Audiences. *Information, Communication & Society*, 12(4):469–487.

[Schermerhorn et al., 2008] Schermerhorn, P., Scheutz, M., and Crowell, C. R. (2008). Robot social presence and gender: Do females view robots differently than males? In *Proc. 2008 Human-Computer Interaction Conf.*, pages 263–270, Amsterdam.

[Schulman, 2013] Schulman, D. (2013). *Embodied Agents for Long-Term Discourse*. PhD thesis, Northeastern University, Boston, MA.

[Segura et al., 2012] Segura, E. M., Cramer, H., Gomes, P. F., Nylander, S., and Paiva, A. (2012). Revive!: reactions to migration between different embodiments when playing with robotic pets. In *Proc. Interaction Design for Children Conf*, pages 88–97, Glasgow.

[Selfhout et al., 2009] Selfhout, M., Branje, S., ter Bogt, T., and Meeus, W. (2009). The role of music preferences in early adolescents' friendship formation and stability. *Journal of Adolescence*, 32(1):95–107.

[Short et al., 1976] Short, J., Williams, E., and Christie, B. (1976). *The social psychology of telecommunications*. Wiley, New York, NY.

[Sidner et al., 2005] Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1):140–164.

[Singer and Feddersen, 2004] Singer, E. and Feddersen, J. (2004). LEMUR: robotic musical instruments. In *Proc. 2004 Multimedia Conf*, pages 184–185, New York City, NY.

[Smith and Fernández-Dols, 1997] Smith, C. A. and Fernández-Dols, J. M. (1997). Componential Approach. In Fernández-Dols, J. M. and Russell, J. A., editors, *The Psychology of Facial Expression*, page 400. Cambridge University Press, Cambridge.

[Soyel and McOwan, 2013] Soyel, H. and McOwan, P. W. (2013). Towards an affect sensitive interactive companion. *Computers & Electrical Engineering*, 39(4):1312–1319.

[Stark, 2011] Stark, A. (2011). *Musicians and Machines: Bridging the Semantic Gap In Live Performance*. PhD thesis, Queen Mary, University of London.

[Stiehl et al., 2009] Stiehl, W. D., Lee, J. K., Breazeal, C., Nalin, M., Morandi, A., and Sanna, A. (2009). The huggable: a platform for research in robotic companions for pediatric care. In *Proc. Interaction Design for Children Conf.*, pages 317–320, Como.

[Sung et al., 2009] Sung, J., Christensen, H. I., and Grinter, R. E. (2009). Robots in the Wild: Understanding Long-Term Use. In *Proc. 2009 Human Robot Interaction Conf.*, pages 45–52, San Diego, CA.

[Suzuki and Umemuro, 2012] Suzuki, D. and Umemuro, H. (2012). Dimensions of people's attitudes toward robots. In *Proc. Human-Computer Interaction Conf.*, pages 249–250, Amsterdam.

[Syrdal et al., 2009] Syrdal, D., Dautenhahn, K., Koay, K., and Walters, M. (2009). The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. In *Proc. Adaptive and Emergent Behaviour and Complex Systems Conf.*, pages 109–115, Edinburgh.

[Syrdal et al., 2008] Syrdal, D., Dautenhahn, K., Walters, M., and Koay, K. (2008). Sharing Spaces with Robots in a Home Scenario–Anthropomorphic Attributions and their Effect on Proxemic Expectations and Evaluations in a Live HRI Trial. In *Proc. AAAI Fall Symposium*, pages 116–123.

[Szafir and Mutlu, 2012] Szafir, D. and Mutlu, B. (2012). Pay attention!: Designing adaptive agents that monitor and improve user engagement. In *Proc. 2012 Human Factors in Computing Systems Conf.*, CHI '12, pages 11–20, New York, NY, USA. ACM.

[Takayama and Pantofaru, 2009] Takayama, L. and Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. In *Proc. 2009 Intelligent Robots and Systems Conf.*, pages 5495–5502.

[Takeda et al., 2007] Takeda, T., Hirata, Y., and Kosuge, K. (2007). Hmm-based error recovery of dance step selection for dance partner robot. In *Proc. Robotics and Automation Conf.*, pages 1768–1773, Roma.

[Tanaka et al., 2007] Tanaka, F., Cicourel, A., and Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, 104(46):17954–17958.

[Theofilis et al., 2013] Theofilis, K., Nehaniv, C., and Dautenhahn, K. (2013). Exploring music as communicative gesture: A drumming implementation for a humanoid robot. In *Proc. 2013 Artificial Life Symposium*, pages 98–104.

[Thompson and Graham, 2005] Thompson, W. and Graham, P. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, (156):203–227.

[Thorne, 2012] Thorne, F. (2012). http://www.felixsmachines.com. Accessed: 2015-05-30.

[Tickle-Degnen, 2006] Tickle-Degnen, L. (2006). Nonverbal behaviour and its functions in the ecosystem of rapport. In Manusov, V. and Patterson, M. L., editors, *The SAGE Handbook of Nonverbal Communiation*. SAGE Publications, Thousand Oaks, CA.

[Tickle-Degnen and Rosenthal, 1990] Tickle-Degnen, L. and Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, 1(4):285–293.

[Tresset and Leymarie, 2006] Tresset, P. and Leymarie, F. F. (2006). AIKON: the Artistic/automatic IKONograph. In *Research Posters: 2006 Computer Graphics and Interactive Techniques Conf.*

[Tryon and Lewis, 2008] Tryon, W. W. and Lewis, C. (2008). An inferential confidence interval method of establishing statistical equivalence that corrects Tryon's (2001) reduction factor. *Psychological methods*, 13(3):272–277.

[Veenstra and Evers, 2011] Veenstra, D. N. and Evers, V. (2011). The Development of an Online Research Tool to Investigate Children's Social Bonds with Robots. In Lamers, M. H. and Verbeek, F. J., editors, *Human Robot Personal Relationships: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 19–26. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Vines et al., 2006] Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, 101(1):80–113.

[Virnyia et al., 2004] Virnyia, Z., Toplb, J., Gcsib, M., dm Miklsia, and Csnyia, V. (2004). Dogs respond appropriately to cues of humans' attentional focus. *Behavioural Processes*, 66(2):161 – 172.

[Wada and Shibata, 2006] Wada, K. and Shibata, T. (2006). Robot therapy in a care house - its sociopsychological and physiological effects on the residents. In *Proc. Robotics and Automation Conf.*, pages 3966–3971, Orlando, FL.

[Wainer et al., 2014] Wainer, J., Robins, B., Amirabdollahian, F., and Dautenhahn, K. (2014). Using the humanoid robot kaspar to autonomously play triadic games and facilitate collaborative play among children with autism. *Autonomous Mental Development, IEEE Transactions on*, 6(3):183–199.

[Walters et al., 2005] Walters, M., Dautenhahn, K., te Boekhorst, R., Koay, K., Kaouri, C., Woods, S., Nehaniv, C., Lee, D., and Werry, I. (2005). The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In *Proc. 2005 Robot and Human Interactive Communication Conf.*, pages 347–352.

[Walters et al., 2009] Walters, M., Koay, K., Syrdal, D., Dautenhahn, K., and Te Boekhorst, R. (2009). Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Proc. of New Frontiers in Human-Robot Interaction : Symposium at AISB09 Convention*, pages 136–143.

[Walters et al., 2011] Walters, M., Oskoei, M., Syrdal, D., and Dautenhahn, K. (2011). A long-term Human-Robot Proxemic study. *Proc. Robot and Human Interactive Communication Conf.*, pages 137–142.

[Weinberg and Driscoll, 2006] Weinberg, G. and Driscoll, S. (2006). Robot-human interaction with an anthropomorphic percussionist. In *Proc. 2006 Human Factors in Computing Systems Conf*, pages 1229–1232, Montreal, QU.

[Weinberg and Driscoll, 2007] Weinberg, G. and Driscoll, S. (2007). The interactive robotic percussionist: new developments in form, mechanics, perception and interaction design. In *Proc. Human-Robot Interaction Conf*, pages 97–104, Arlington, VA.

[Weinberg et al., 2009] Weinberg, G., Raman, A., and Mallikarjuna, T. (2009). Interactive jamming with Shimon: a social robotic musician. In *Proc. 2009 Human Robot Interaction Conf.*, pages 233–234, San Diego, CA.

[Weiss, 1974] Weiss, R. (1974). The Provisions of Social Relationships. In Rubin, Z., editor, *Doing Unto Others*, pages 17–26. Prentice Hall, Englewood Cliffs, NJ.

[Winkler, 2000] Winkler, T. (2000). *Interactive Music: Techniques for Composing with Max*. MIT Press, Cambridge, MA.

[Wiseman, 1986] Wiseman, J. P. (1986). Friendship: Bonds and Binds in a Voluntary Relationship. *Journal of Social and Personal Relationships*, 3(2):191–211.

[Wright, 1984] Wright, P. H. (1984). Self-Referent Motivation and the Intrinsic Quality of Friendship. *Journal of Social and Personal Relationships*, 1(1):115–130.

[Xenakis, 1992] Xenakis, I. (1992). *Formalized Music: Thought and Mathematics in Composition* . Indiana University Press, Bloomington, IN.

[Yuan et al., 2006] Yuan, J., Liberman, M., and Cieri, C. (2006). Towards an Integrated Understanding of Speaking Rate in Conversation. In *Proc. 2006 Spoken Language Processing Conf.*, pages 541–544, Pittsburgh, PA.

[Zoll et al., 2010] Zoll, C., Diruf, M., and Spielhagen, C. (2010). Preferences for Robotic Home Companions – A Conjoint Analysis. In *Proc. 2010 Intelligent Systems and Agents Conf.*, pages 164–166, Freiburg.